

# C&S SIG

---

## ***ESTRATÉGIA MULTI-TEMPORAL PARA PRODUÇÃO AUTOMÁTICA DE CARTOGRAFIA DE OCUPAÇÃO DO SOLO COM IMAGENS AWIFS***

---

Hugo Alexandre Gomes da Costa

---

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência e Sistemas de Informação Geográfica

---

Instituto Superior de Estatística e Gestão de Informação  
da Universidade Nova de Lisboa

# **ESTRATÉGIA MULTI-TEMPORAL PARA PRODUÇÃO AUTOMÁTICA DE CARTOGRAFIA DE OCUPAÇÃO DO SOLO COM IMAGENS AWiFS**

Dissertação orientada por

Professor Doutor Mário Caetano

e

Professor Doutor Fernando Bação

Novembro de 2008

## **AGRADECIMENTOS**

A elaboração da presente dissertação teve o contributo de diversas pessoas a quem não posso deixar de agradecer.

Um agradecimento especial ao Professor Doutor Mário Caetano, pela oportunidade e confiança depositada em mim e pela orientação científica, colaboração e acompanhamento no desenvolvimento deste trabalho. Este agradecimento é também extensível ao Instituto Geográfico Português pelas condições de trabalho indispensáveis que proporcionou.

Ao Professor Doutor Fernando Bação, um agradecimento pela disponibilidade, sugestões e ideias, e ainda pela simpatia.

Agradeço também aos meus colegas do Grupo de Detecção Remota, pelo óptimo ambiente de trabalho, amizade, troca de ideias e críticas que em muito enriqueceram este trabalho, em especial ao António Araújo e Hugo Carrão.

Por fim, o meu (incomparável) maior agradecimento aos meus pais, pelo amor, apoio e conselhos que sempre me dão.

# **ESTRATÉGIA MULTI-TEMPORAL PARA PRODUÇÃO AUTOMÁTICA DE CARTOGRAFIA DE OCUPAÇÃO DO SOLO COM IMAGENS AWiFS**

## **RESUMO**

A produção de cartografia de ocupação do solo por métodos tradicionais (interpretação visual de imagens aéreas e de satélite) é um processo com custos que inviabilizam a sua rápida produção e frequente actualização. Esta realidade é bastante lesante para diversas aplicações que necessitam deste género de informação actualizada. Uma solução viável para fazer face à problemática da desactualização da cartografia de ocupação do solo é a criação de um produto menos detalhado do que a cartografia produzida pelos programas operacionais já existente. Isto permite que o novo produto seja mais barato e possa ser produzido por métodos automáticos, realizável regularmente, que ofereça informação actual e seja útil para diversas aplicações. Neste sentido, foi explorado um conjunto de dados intra-anual composto por três imagens AWiFS (Abril, Julho e Outubro de 2006) para avaliar as imagens e métodos automáticos de produção que possam derivar cartografia de ocupação do solo de Portugal Continental no âmbito de um novo programa operacional anual. Foram testados dois classificadores: um classificador paramétrico convencional (classificador da máxima verosimilhança) e um classificador não paramétrico (árvores de decisão). Foram desenhados vários testes para avaliar a melhor abordagem de classificação, o efeito do tamanho da amostragem de treino na exactidão da classificação e a aptidão das imagens AWiFS para derivar a cartografia pretendida. A exactidão global dos mapas produzidos variou à volta de 60% e 72% com uma nomenclatura de 15 e 10 classes de ocupação do solo, respectivamente. Os resultados obtidos sugerem que as imagens AWiFS apresentam algumas limitações para derivar cartografia de ocupação do solo de Portugal Continental. Ao nível da metodologia, os resultados revelam que as árvores de decisão são um classificador vantajoso relativamente ao classificador da máxima verosimilhança porque permite uma fase de treino mais rápida e com um menor esforço de amostragem sem perdas de exactidão temática.

# **AUTOMATIC MULTI-TEMPORAL STRATEGY TO LAND COVER CARTOGRAPHY PRODUCTION WITH AWiFS IMAGERY**

## **ABSTRACT**

Land cover map production through usual methods (visual interpretation of aerial and satellite images) is an expensive process that makes its quick production and frequent updating unfeasible. This fact is very prejudicial to several applications that demand for this type of information updated. A possible solution to face the outdating problem of land cover maps is to create a new product less detailed than the cartography produced by existent operational programs. This allows the new product to be cheaper and produced with automatic methods, updated frequently and useful for several purposes. Therefore, three AWiFS scenes from three intra-annual time periods (April, July and October 2006) were explored to assess the images and automatic methods of production that can derive land cover maps of mainland Portugal in an annual operational basis. Two different classifiers were tested: a more conventional parametric classifier, the maximum likelihood classifier, and a nonparametric classifier, a decision tree. Several tests were designed to evaluate the most suitable classification approach, training sample size effect on the classification accuracy and the images' ability for land cover mapping of mainland Portugal. Overall accuracy values, of the land cover maps were around 60% and 72% with a 15 and 10 class nomenclature, respectively. The quality of the results were moderate regardless the classification approach and training sample characteristics which suggests that AWiFS images have limitations for land cover mapping the Portuguese landscape. Results also reveal that decision tree classifier is advantageous relatively to maximum classifier likelihood because allows quicker training and a smaller sampling effort without accuracy losses.

## **PALAVRAS-CHAVE**

Árvores de decisão

AWiFS

Cartografia de ocupação do solo

Classificador da máxima verosimilhança

Portugal

## **KEYWORDS**

Decision trees

AWiFS

Land cover map

Maximum likelihood classifier

Portugal

## ACRÓNIMOS

**AD** – Árvores de Decisão

**AVHRR** – Advanced Very High Resolution Radiometer

**AWiFS** – Advanced Wide Field Sensor

**ASAR** - Advanced Synthetic Aperture Radar

**CLC2000** – CORINE Land Cover

**CLC2000** – CORINE Land Cover 2000

**CMV** – Classificador da Máxima Verosimilhança

**CORINE** – Co-ORdination of INformation on the Environment

**COS** – Cartografia de Ocupação do Solo

**COS'90** – Cartografia de Ocupação do Solo de 1990

**COSMIC** – Cartografia multi-escala de ocupação do solo

**CP** – Componentes Principais

**DGRF** – Direcção-Geral dos Recursos Florestais

**EG** – Exactidão Global

**EP** – Exactidão do Produtor

**EU** – Exactidão do Utilizador

**FAO** – Food and Agriculture Organization

**GDR** – Grupo de Detecção Remota

**IA** – Instituto do Ambiente

**IGBP** – International Geosphere-Biosphere Programme

**IGP** – Instituto Geográfico Português

**IVV** – Instituto da Vinha e do Vinho

**LCCS** – Land Cover Classification System

**LISS** – Light Intensity Scanning System

**MERIS** – Medium Resolution Imaging Spectrometer

**MODIS** – Moderate Resolution Imaging Spectroradiometer

**MDE** – Modelo Digital de Elevação

**NR-Census** – Natural Resource Census

**NDVI** – Normalized Difference Vegetation Index

**NLCD** – National Land-Cover Database

**SPOT** – Systeme Pour l’Observation de la Terre

**SPREADII** – Spatial REallocation of Aggregated Data

**WiFS** – Wide Field Sensor



## ÍNDICE DO TEXTO

<b>AGRADECIMENTOS</b> .....	<b>iii</b>
<b>RESUMO</b> .....	<b>iv</b>
<b>ABSTRACT</b> .....	<b>v</b>
<b>PALAVRAS-CHAVE</b> .....	<b>vi</b>
<b>KEYWORDS</b> .....	<b>vi</b>
<b>ACRÓNIMOS</b> .....	<b>vii</b>
<b>ÍNDICE DE TABELAS</b> .....	<b>xi</b>
<b>ÍNDICE DE FIGURAS</b> .....	<b>xiii</b>
<b>1. Introdução</b> .....	<b>1</b>
1.1. Enquadramento .....	1
1.2. Objectivos .....	2
1.3. Organização da dissertação .....	3
<b>2. Cartografia de ocupação do solo e detecção remota</b> .....	<b>4</b>
2.1. Produção de cartografia de ocupação do solo .....	4
2.1.1. Nomenclaturas de ocupação do solo .....	4
2.1.2. Fontes de dados de detecção remota .....	6
2.1.3. Métodos de extracção de informação de dados de detecção remota.....	8
2.1.4. Avaliação da qualidade da cartografia.....	10
2.2. Programas operacionais de cartografia de ocupação do solo .....	12
2.2.1. Mapa de ocupação do solo da Grã-Bretanha.....	14
2.2.2. CORINE Land Cover 2000 e Mapa de Ocupação do Solo suecos .....	15
2.2.3. Programa <i>Australian Collaborative Land Use Mapping</i> .....	16
2.2.4. Projecto <i>National Land-Cover Database</i> (NLCD 2001) .....	17
2.2.5. Mapa de ocupação do solo do programa indiano NR-Census.....	19
<b>3. Metodologia</b> .....	<b>21</b>
3.1. Estratificação geográfica .....	25
3.2. Pré-processamento das imagens.....	28
3.3. Nomenclatura de ocupação do solo .....	29
3.4. Classificação assistida .....	30
3.4.1. Os classificadores da máxima verosimilhança e árvores de decisão .....	31
3.4.2. Amostragem de treino.....	33

3.4.3. Testes de classificação.....	36
3.5. Avaliação da exactidão do mapa.....	43
<b>4. Resultados e discussão .....</b>	<b>48</b>
4.1. Exactidão global dos testes de classificação .....	48
4.2. Variáveis dos testes de classificação .....	54
4.3. Matrizes de confusão dos melhores testes ao nível do <i>pixel</i> .....	56
4.4. Comparação dos melhores testes ao nível do <i>pixel</i> com o CLC2000 .....	69
4.5. Imagens AWiFS.....	70
4.6. Críticas e perspectivas futuras .....	72
<b>5. Conclusões .....</b>	<b>75</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>77</b>
<b>ANEXOS .....</b>	<b>81</b>
1. Imagens AWiFS (composição colorida 342).....	82
2. Nomenclaturas de ocupação do solo dos programas operacionais de cartografia de ocupação do solo .....	83
3. Nomenclatura CORINE Land Cover .....	86
4. Nomenclatura LANDEO e nomenclatura adoptada .....	87
5. Exactidão global e classes de cada fase de classificação do teste 15.....	89
6. Comparação gráfica entre os melhores mapas obtidos com o CMV e AD .....	90

## ÍNDICE DE TABELAS

Tabela 1 – Projectos operacionais de cartografia de ocupação do solo que usam métodos automáticos .....	13
Tabela 2 – Características das imagens AWiFS (Fonte: NRSA, 2005).....	21
Tabela 3 – Parâmetros de aquisição das imagens AWiFS.....	22
Tabela 4 – Informação auxiliar .....	23
Tabela 5 – Níveis de estratificação geográfica e suas características gerais .....	27
Tabela 6 – Sistemas de representação das imagens antes e após a ortorectificação .....	28
Tabela 7 – Erro médio quadrático (RMS) da ortorectificação das imagens AWiFS.....	28
Tabela 8 – Nomenclatura de ocupação do solo .....	30
Tabela 9 – Classes espectrais passíveis de serem usadas na fase de treino .....	35
Tabela 10 – Características dos testes de classificação .....	37
Tabela 11 – Avaliações permitidas através da comparação de testes de classificação .....	38
Tabela 12 – Probabilidades <i>a priori</i> usadas no teste 8 e calculadas a partir do CLC2000 ...	39
Tabela 13 – Nomenclaturas de ocupação do solo com 15, 10 e 6 classes.....	40
Tabela 14 – Nomenclaturas de ocupação do solo menos detalhadas usadas no teste 14 ..	41
Tabela 15 – Esquema de classificação por fases do teste 14.....	42
Tabela 16 – Resultados e características dos testes de classificação.....	48
Tabela 17 – Matriz de confusão do teste 6 com o classificador da máxima verosimilhança	62
Tabela 18 – Matriz de confusão do teste 15 com árvores de decisão .....	66
Tabela 19 – Breves considerações acerca das classes e classificadores com base nas matrizes de confusão e nos gráficos do Anexo 6 .....	67
Tabela 20 – Valores de exactidão global (EG) não ponderada e ponderada do teste 6 com CMV .....	68
Tabela 21 – Comparação entre o mapa do teste 6 com o CMV e o mapa CLC2000 .....	69
Tabela 22 – Comparação entre o mapa do teste 15 com árvores de decisão e o mapa CLC2000 .....	70
Tabela 2.1 – Nomenclatura do Mapa de ocupação do solo da Grã-Bretanha .....	83
Tabela 2.2 – Classes florestais do Mapa de Ocupação do Solo sueco e usadas na classificação automática .....	83

Tabela 2.3 – Nomenclatura do mapa do programa <i>Australian Collaborative Land Use Mapping</i> .....	84
Tabela 2.4 – Nomenclatura do projecto NLCD 2001 .....	84
Tabela 2.5 – Nomenclatura do Mapa de ocupação do solo do programa indiano NR-Census .....	85
Tabela 3.1 – Nomenclatura do CORINE Land Cover .....	86
Tabela 4.1 – Nomenclatura LANDEO .....	87
Tabela 4.2 – Classes LANDEO adoptadas na presente dissertação (correspondentes códigos e nomes das classes da nomenclatura entre parêntesis) e sua descrição baseada no LCCS .....	88
Tabela 5.1 – Exactidão global e classes de cada fase de classificação do teste 15 .....	89

## ÍNDICE DE FIGURAS

Figura 1 – Matriz de confusão com q classes de ocupação do solo e índices de qualidade do mapa (Fonte: Carrão <i>et al.</i> 2007).....	12
Figura 2 – Área coberta pelas seis imagens AWiFS .....	21
Figura 3 – Exemplo de diferentes estados espectrais ao longo do ano de áreas agrícolas irrigadas.....	23
Figura 4 – Representação geral da metodologia.....	24
Figura 5 – Unidades de paisagem (Fonte: DGOTDU, 2004).....	26
Figura 6 – Estratificação de Portugal Continental.....	27
Figura 7 – Amostra de treino.....	34
Figura 8 – Classificação por fases desenvolvida no teste de classificação 15 .....	43
Figura 9 – Amostra de validação .....	44
Figura 10 – Evolução da exactidão temática da estratégia de classificação por fases do teste 15 .....	52
Figura 11 – Exemplo da segmentação das imagens AWiFS .....	53
Figura 12 – Mapa de ocupação do solo produzido no teste 6 com o classificador da máxima verosimilhança .....	58
Figura 13 – Exemplo de possível confusão entre Áreas agro-florestais e Cultura de sequeiro presente no sobcoberto .....	59
Figura 14 – Mapa de ocupação do solo produzido no teste 15 com árvores de decisão .....	63
Figura 17 – Largura das bandas dos sensores AWiFS e MERIS.....	72
Figura 18 – Comparação entre o aspecto visual das classificações ao nível do <i>pixel</i> e híbrida <i>pixel/objecto</i> .....	74
Figura 6.1 – Exemplo de interpretação dos gráficos do Anexo 6 (classe 1 com o CMV do teste 6) .....	90

# 1. Introdução

## 1.1. Enquadramento

A produção de cartografia de ocupação do solo é de grande importância estratégica em diversas áreas. A sua aplicação é transversal e apresenta-se fundamental no planeamento e ordenamento do território (Foody e Mathur, 2004; Homer *et al.*, 2007), modelação de fenómenos ambientais (Friedl *et al.*, 2002; Justice *et al.*, 1998), estudo e análise espacial de sistemas biofísicos (Stehman e Czaplewski, 1998; Justice *et al.*, 1998; Homer *et al.*, 2007) e/ou humanos, monitorização ambiental (Homer *et al.*, 2007), detecção de alterações, etc.

Em Portugal existe cartografia de ocupação do solo produzida em diferentes programas e com especificações próprias (data de referência, unidade mínima cartográfica, etc.). Os mais conhecidos são os programas de produção da Carta de Ocupação do Solo e CORINE Land Cover (CLC) cujos produtos de 1990 (COS'90) e 2000 (CLC2000) continuam a ser usados como informação temática de referência em inúmeras aplicações (Caetano *et al.*, 2008).

Como toda a cartografia temática, a informação acerca da ocupação do solo de um território refere-se a um determinado momento temporal. Isto quer dizer que a informação representada pela cartografia se desactualiza com o passar do tempo, até que deixa de ser útil quando se deseja ter informação actualizada. A disponibilidade de informação actualizada é um imperativo para diversos estudos e aplicações que se tornam reféns da referência temporal da informação cartográfica existente.

A produção de cartografia de ocupação do solo à escala nacional não é um objectivo fácil e rápido de atingir. No seu planeamento tem que se atender aos dados de referência que servirão para a produção da cartografia (imagens de satélite, imagens aéreas, etc.), prever os objectivos genéricos das futuras utilizações da cartografia, definir o detalhe geométrico e temático de representação, a metodologia a adoptar, etc. Por todos estes factores, a produção regular de cartografia de ocupação do solo comporta elevados custos, o que leva a que muitas vezes não exista informação actualizada.

A utilidade da realização regular, de preferência anual, de cartografia como a COS'90 ou o CLC2000 é inquestionável, mas igualmente impossível, devido aos custos referidos anteriormente. Estes são projectos de escala detalhada, assim como a nomenclatura adoptada, realizados por interpretação visual de imagens aéreas e satélite, respectivamente, o que exige um elevado número de pessoas e tempo para a sua realização. Estas condicionantes fazem com que haja um défice de informação no período de tempo entre projectos de produção de cartografia de ocupação do solo. Segundo uma revisão de programas operacionais de cartografia de ocupação de várias regiões do mundo,

realizada por Caetano *et al.* (2002), verifica-se que a média de actualização de produtos deste género é de 10-15 anos.

Para fazer face ao problema da actualidade da informação de ocupação do solo, impõe-se pensar em estratégias que complementem a informação oferecida pelos grandes programas de escala nacional já existentes (COS e CLC). Como não é possível a realização de cartografia de ocupação do solo detalhada em menores períodos de tempo, uma solução viável é o desenvolvimento de um novo produto menos detalhado (mais barato e rápido de produzir), realizável regularmente, que ofereça informação actual e seja útil para diversas aplicações. Neste contexto surge o projecto COSMIC–Cartografia multi-escala de ocupação do solo (doravante apenas designado por COSMIC).

O projecto COSMIC é um projecto de produção de cartografia de ocupação do solo da responsabilidade do Grupo de Detecção Remota (GDR) do Instituto Geográfico Português (IGP) e financiado pelo Programa de Investimentos e Despesas de Desenvolvimento da Administração Central (PIDDAC). Tem a pretensão de desenvolver um produto intermédio entre os programas já existentes de forma a mitigar a desactualização da informação cartográfica de ocupação do solo em Portugal Continental. O desenvolvimento do projecto tem estabelecido pontos de contacto com o projecto LANDEO, financiado pela Fundação para a Ciência e Tecnologia (FCT) e Agência Espacial Europeia (ESA). O seu objectivo é desenvolver metodologias integradas para explorar dados ópticos e ASAR obtidos pelo satélite ENVISAT para caracterização da ocupação de solo. Embora o projecto COSMIC tenha um objectivo mais operacional, os contactos com um projecto de I&D como o LANDEO têm sido bastante benéficos no seu desenvolvimento. A presente dissertação apresenta o trabalho desenvolvido no âmbito do projecto COSMIC, no qual se tem investigado metodologias automáticas de classificação de imagens de satélite para produção de cartografia de ocupação do solo. O trabalho desenvolvido e aqui apresentado relaciona-se com uma área de estudo localizada nas bacias dos rios Tejo e Sado, criteriosamente seleccionada para a realização de testes e desenvolver a metodologia a aplicar em todo o território continental.

## **1.2. Objectivos**

Na sequência das carências identificadas em informação actualizada acerca da ocupação do solo, foi definido um objectivo genérico. Este consiste do desenvolvimento de uma metodologia de classificação automática de imagens de satélite obtidas pelo sensor *Advanced Wide Field Sensor* (AWiFS), exequível para a produção de um novo mapa de ocupação do solo intermédio entre os mapas dos programas já existentes. Este objectivo genérico incorpora os seguintes objectivos específicos:

- i) Explorar o potencial das imagens AWiFS para discriminar diferentes classes de ocupação do solo;

- ii) Desenvolver uma nomenclatura de ocupação do solo ajustada às potencialidades do conjunto de imagens usado e à paisagem portuguesa;
- iii) Desenvolver e testar uma metodologia exequível anualmente;
- iv) Desenvolver e testar uma metodologia expedita e que não necessite em elevados recursos financeiros;
- v) Produzir um mapa de ocupação do solo menos detalhado que os mapas produzidos nos programas de referência já existentes, mas útil para diversas aplicações exigentes em informação actualizada.

### **1.3. Organização da dissertação**

Após uma introdução ao tema, na presente dissertação é feita uma breve revisão das fases por que passam os programas de produção de cartografia de ocupação do solo (ponto 2.1) e são revistos, de uma forma breve, programas operacionais de ocupação do solo baseados em métodos automáticos (2.2). No ponto 3 é apresentada a metodologia desenvolvida, nomeadamente, a estratificação geográfica (3.1), o pré processamento das imagens (3.2), a definição da nomenclatura de ocupação do solo (3.3), a classificação assistida das imagens de satélite (3.4) e a avaliação da exactidão (3.5). Os resultados obtidos são apresentados e discutidos no ponto 4, assim como referidas críticas e perspectivas futuras ao trabalho desenvolvido. No ponto 5 são apresentadas as conclusões.



## **2. Cartografia de ocupação do solo e detecção remota**

### **2.1. Produção de cartografia de ocupação do solo**

A detecção remota pode ser definida como a ciência de aquisição, processamento e interpretação de imagens adquiridas por avião e satélite que gravam a interacção entre a matéria e a energia electromagnética (Sabins, 1987). É uma ciência com múltiplas aplicações, uma das mais comuns, a derivação de cartografia de ocupação do solo (Foody, 2002). Apesar da considerável quantidade de projectos de cartografia de ocupação do solo desenvolvidos por todo o mundo, não existe uma metodologia padrão para esta tarefa e continua a ser uma área de investigação e novos desenvolvimentos. Ainda assim, de uma forma geral, todos os projectos desta natureza passam por definir a nomenclatura de ocupação do solo, seleccionar dados obtidos por detecção remota, seleccionar métodos de extracção da informação incorporada nas imagens e avaliar a qualidade da cartografia produzida.

#### **2.1.1. Nomenclaturas de ocupação do solo**

Qualquer processo de produção de cartografia temática tem, na sua fase inicial, que definir a nomenclatura do tema em questão. A definição da nomenclatura deve ser ponderada em função dos objectivos da cartografia, das suas características técnicas (e.g. unidade mínima cartográfica), do método de produção e deverá respeitar critérios de coerência e clareza.

Existem várias nomenclaturas de ocupação do solo de referência a nível internacional. Os mais citados e difundidos são a do projecto CORINE Land Cover, do *United States Geological Survey*, do *International Geosphere-Biosphere Programme* (IGBP) e da *Food and Agriculture Organization* (FAO) das Nações Unidas. Ao nível nacional, recentemente foi desenvolvida a nomenclatura LANDEO.

A nomenclatura CORINE Land Cover surge no âmbito do programa Co-ORDination of INformation on the Environment (CORINE), estabelecido pela Comissão Europeia e desenvolvido com o objectivo de obter informação ambiental comparável a nível europeu para servir de base à definição e monitorização de políticas europeias de uma forma eficiente (Montanarella *et al.*, 2003). As classes foram definidas de forma a classificar a ocupação do solo na União Europeia à escala 1:100000 através de interpretação visual de imagens de satélite (Landsat e SPOT) e informação auxiliar (Bossard *et al.*, 2000). A nomenclatura inclui 44 classes agrupadas em três níveis hierárquicos. O primeiro nível agrupa cinco classes de carácter generalista (Territórios artificializados, Área agrícolas, Florestas e meio semi-naturais, Zonas húmidas e Massas de Água), o segundo nível agrupa 15 classes e o terceiro nível, o mais detalhado, agrupa 44 classes (Anexo 3).

A nomenclatura do *United States Geological Survey*, conhecida por *USGS Land use/Land cover Classification system*, assenta na nomenclatura definida por Anderson *et al.* (1976). Esta nomenclatura é hierárquica, de quatro níveis e foi pensada para ser usada com imagens de satélite e imagens aéreas. O primeiro nível engloba nove classes generalistas (*Water, developed, barren, forest upland, shrubland, non-nature woody, herbaceous upland natural/semi-natural vegetation, herbaceous planted/cultivated* e *wetlands*). O segundo nível é composto por 24 classes representáveis a uma escala inferior a 1:80000. O terceiro e quarto níveis são compostos por classes muito detalhadas, identificadas com imagens de grande ou muito grande resolução espacial e podem ser usadas a escalas superiores a 1:80000 e 1:20000, respectivamente.

A nomenclatura IGBP foi desenvolvida para ser usada em cartografia de ocupação do solo derivada de imagens do sensor *Advanced Very High Resolution Radiometer* (AVHRR), cuja necessidade se fazia sentir em várias iniciativas do *International Geosphere-Biosphere Programme* (Townshend *et al.* 1994 in Loveland e Belward, 1997). A nomenclatura é composta por 17 classes, concebidas para classificar a totalidade da superfície terrestre e de forma a serem interpretáveis através de dados de detecção remota com pelo menos 1Km de resolução espacial ou com dados de campo. A definição das classes compreende a filosofia de independência de tipos de clima e componentes da cobertura vegetal apresentada por Running *et al.* (1994), mas modificada para serem compatíveis com as nomenclaturas usadas em modelação ambiental e para fornecer, quando possível, implicações acerca do uso do solo e para representar mosaicos (Loveland e Belward, 1997). As 17 classes dividem-se em 11 classes de vegetação natural, três classes desenvolvidas ou mosaico e duas classes de não vegetação.

Devido à existência de várias nomenclaturas, que assentam em princípios diferentes e são, por isso, incompatíveis, coloca-se um problema quando é necessário usar diferentes produtos cartográficos. Esta necessidade levou a FAO a desenvolver uma iniciativa com o objectivo de responder à necessidade de uma abordagem estandardizada, de onde surgiu o *Land Cover Classification System* (LCCS). Este é um sistema de classificação normalizada *a priori*, comparável a sistemas existentes, aplicável a qualquer escala, em todo o mundo e independente do método de aquisição de dados ou da construção de cartografia. O sistema inclui duas fases: uma fase dicotómica e uma fase modelar hierárquica. Na primeira fase são distinguidas oito classes principais (*Cultivated terrestrial, Natural/semi-natural terrestrial, Cultivated aquatic, Natural/semi-natural aquatic, Built-up & associated areas, Bare areas, Artificial water bodies* e *Inland water*). Na segunda fase, a criação de cada classe é mediante a combinação de um conjunto predefinido de classes puras (Gregorio e Jansen, 1996).

A inexistência de uma nomenclatura *standard* para mapas de ocupação de solo levou à necessidade de desenvolver a nomenclatura LANDEO (Araújo e Caetano, 2006). Assim, o

GDR desenvolveu esforços para estabelecer uma nomenclatura simples, clara, com limites de classes bem definidos, que seja aplicável na produção de mapas de ocupação de solo a partir de metodologias de classificação automática e a diferentes escalas de análise. O desenvolvimento da nomenclatura baseou-se no LCCS da FAO e teve em consideração as especificidades do território português e as nomenclaturas de ocupação do solo já existentes (de forma a serem comparáveis) (Araújo e Gaetano, 2006). É uma nomenclatura hierárquica de cinco níveis. O primeiro nível é composto por sete classes generalistas (Áreas artificiais, Agricultura, Vegetação natural e semi-natural, Mosaicos, Zonas húmidas permanentes, Solo nu e Água). Os restantes níveis consistem na separação de cada classe do nível anterior em classes cada vez mais detalhadas, num total de 37 classes no quinto nível (Anexo 4). A nomenclatura LANDEO foi usada como base para a definição da nomenclatura da cartografia derivada das imagens AWiFS no âmbito do projecto COSMIC, apresentada no ponto 3.3.

### **2.1.2. Fontes de dados de detecção remota**

A nível nacional, as fontes de dados exequíveis de serem usadas na produção de cartografia de ocupação do solo são imagens de satélite ou aéreas. Apesar de se assistir a uma aproximação entre estes dois tipos de imagens devido à evolução tecnológica da sua obtenção, as imagens de satélite são o recurso mais vantajoso face aos objectivos definidos.

Vários factores contribuem para esta situação. As imagens aéreas, embora permitam a produção de cartografia bastante precisa, incorporam custos bastantes elevados relativos a três aspectos: aquisição, interpretação e tempo de execução (Lillesand e Kiefer, 1994). Relativamente ao primeiro aspecto, para além de se revelar dispendioso, este tipo de imagens exige o planeamento prévio dos voos, o que limita a disponibilidade de dados em datas anteriores. Os segundo e terceiro aspectos estão relacionados com o método de extracção da informação das imagens. As imagens aéreas estão tradicionalmente mais associadas a métodos manuais de interpretação visual, realizados por equipas de intérpretes, que são por natureza morosos e dispendiosos. Pelo contrário, as imagens de satélite, obtidas sistematicamente segundo as características orbitais do satélite, permitem a obtenção de imagens sem planeamento prévio, em grande quantidade e a baixo preço. No que se refere aos métodos de extracção da informação das imagens, estas estão mais associadas a métodos automáticos de classificação, mais baratos e rápidos de executar (Gaetano *et al.*, 2002).

Outra característica distintiva das imagens aéreas e de satélite é a área que cobrem. Normalmente, no primeiro caso, uma imagem cobre algumas dezenas de quilómetros quadrados e no segundo caso, cada imagem pode cobrir milhares de quilómetros quadrados. Esta característica determina o número de imagens necessário para se deter

informação que cubra todo o território a cartografar, o que tem grande influência no processo de uma classificação por métodos automáticos. O uso de diversas imagens que abrangem grandes áreas proporciona uma variedade de padrões espectrais entre imagens relacionados com a fisiografia, fenologia, condições de iluminação, atmosféricas e do próprio sensor (Homer *et al.*, 2004). Assim, num processo automático de classificação de imagens, é favorável deter todo o território num menor número de imagens para reduzir a variabilidade espectral não relacionada com as características espectrais das classes de ocupação do solo de interesse.

Em Portugal, têm sido exploradas imagens de satélite adquiridas por vários sensores, nomeadamente por parte do GDR do IGP: Landsat, MERIS, MODIS, Ikonos, etc. No entanto, apenas as imagens Landsat foram usadas na criação de cartografia no âmbito de um programa operacional (CORINE Land Cover) e mais recentemente, imagens MERIS no âmbito do projecto COSMIC (Carrão *et al.*, no prelo). Ao nível internacional, a maioria dos produtos cartográficos são gerados com imagens Landsat TM, provavelmente em consequência do seu reduzido preço (Caetano *et al.*, 2002).

Ainda em relação aos dados de base usados na produção de cartografia de ocupação do solo, pode ser favorável o uso de informação auxiliar, como derivados de modelos digitais de elevação (MDE). Se o método de produção for manual (interpretação visual), a informação auxiliar pode ser consultada ao longo do processo como forma de eliminar dúvidas de interpretação. Se o método for automático, a informação auxiliar pode ser combinada com as imagens através de uma estratificação pré-classificação (divisão da área a cartografar com base na informação auxiliar e posterior classificação de cada estrato de forma independente), de melhoramento pós-classificação (em que a informação auxiliar é usada para resolução de dúvidas) e da própria integração da informação auxiliar como bandas adicionais, juntamente com as imagens, na fase da classificação (Caetano *et al.*, 2002).

No presente trabalho foram usadas as imagens do sensor AWiFS, montado no satélite IRS-P6 (também conhecido por Resourcesat). Estas imagens têm sido relativamente pouco usadas para produção de mapas de ocupação do solo e, no caso do território português, nunca o foram. Apesar da reduzida experiência em usar estas imagens, as suas características técnicas (Tabela 2) apresentam um potencial para a discriminação espectral da paisagem portuguesa que merece ser estudado. A resolução espacial é adequada para a produção de cartografia no âmbito dos objectivos definidos; as quatro bandas registam radiação electromagnética da mesma zona do espectro electromagnético que outros sensores que são usados com sucesso para objectivos semelhantes; e a resolução radiométrica assegura um elevado detalhe de resposta à energia captada pelo sensor, o que possibilita uma melhor distinção entre diferentes classes de ocupação do solo com respostas espectrais semelhantes. Além destes factores, as imagens AWiFS podem ser

adquiridas por um preço relativamente baixo, o que as tornam merecedoras de serem avaliadas no âmbito do projecto COSMIC.

### **2.1.3. Métodos de extracção de informação de dados de detecção remota**

A extracção de informação de dados de detecção remota pode ser realizada através de interpretação visual ou classificação automática. Pode-se ainda usar um método híbrido, designado semi-automático, que normalmente consiste numa classificação automática seguida de correcção de erros por interpretação visual.

Por interpretação visual entende-se examinar imagens de detecção remota com o propósito de identificar objectos na superfície terrestre e disponibilizar esta informação (Lillesand e Kiefer, 1994). No caso específico da produção de cartografia de ocupação do solo, consiste na identificação das classes de ocupação do solo e na sua delimitação sob a imagem, actualmente realizada inteiramente com recurso ao computador. No processo interpretativo, as classes são identificadas através de atributos espectrais e geométricos como a cor, tamanho, forma, textura, padrões espaciais, localização e relações com outras classes. É um método que permite a derivação de cartografia tematicamente bastante exacta devido às excelentes capacidades humanas em avaliar qualitativamente padrões espaciais em imagens. Por outro lado, é bastante demorado, dispendioso e sensível à experiência do intérprete. As características espectrais não são sempre avaliadas totalmente por interpretação visual devido às capacidades humanas em distinguir tonalidades e à limitação de analisar várias imagens em simultâneo (Lillesand e Kiefer, 1994).

Através da classificação de imagens, a interpretação visual é substituída por técnicas quantitativas (algoritmos) para automatizar a identificação das classes de ocupação do solo. Existem diferentes métodos que se podem categorizar em (Lillesand e Kiefer, 1994):

- i) Reconhecimento de padrões espectrais;
- ii) Reconhecimento de padrões espaciais;
- iii) Reconhecimento de padrões temporais.

Reconhecimento de padrões espectrais é um termo atribuído aos métodos que classificação ao nível do *pixel* onde os dados espectrais fornecidos pelas imagens são usados como base numérica para a classificação. O pressuposto é que as diferentes classes estão associadas a diferentes combinações de números digitais nas imagens devido às respectivas reflectâncias espectrais inerentes.

Por outro lado, reconhecimento de padrões espaciais são métodos de classificação que se baseiam nas relações espaciais entre os *pixels*, como a textura, tamanho, distância, forma, repetição e contexto. Este tipo de classificação tenta replicar a estratégia e as capacidades

humanas de classificar por interpretação visual, e por isso, mais complexo e intenso computacionalmente.

O reconhecimento de padrões temporais usa o tempo como uma ajuda à classificação. Neste caso, a classificação é realizada com base em dados recolhidos em diferentes períodos de tempo de forma a aumentar a informação acerca de cada classe. A informação temporal pode ser crucial para o sucesso de uma classificação porque, por vezes, os dados espectrais e espaciais são insuficientes para distinguir classes que apenas se distinguem em determinados períodos de tempo.

Apesar de o factor tempo ser reconhecido como determinante na classificação automática de imagens, pode-se atender apenas aos padrões espectrais e espaciais das imagens. Isto acontece porque os padrões temporais são incorporados no processo de classificação através da adição de mais dados espectrais e espaciais (adição de mais imagens).

Outra forma de categorizar os diferentes métodos de classificação está relacionada com a unidade espacial de análise e consiste em 3 tipos (Caetano *et al.*, 2002):

- i) Classificação ao nível do *pixel*;
- ii) Classificação ao nível do *pixel* combinada com informação contextual;
- iii) Análise de imagens orientada por objectos.

Classificação ao nível do *pixel* é o método automático convencional de extrair informação de imagens. Cada *pixel* é classificado com base na informação espectral que incorpora e independentemente da classificação atribuída aos seus vizinhos.

A classificação ao nível do *pixel* combinada com análise contextual consiste numa análise adicional à classificação ao nível do *pixel* com vista a identificar classes que não podem ser identificadas somente ao nível do *pixel*, como por exemplo, espaços verdes urbanos ou sistemas parcelares complexos. A imagem já classificada é então sujeita a uma análise que se baseia na lógica utilizada da interpretação visual, em que o analista faz uma análise do contexto para identificar vários tipos de classes (Caetano *et al.*, 2002).

Os problemas com a identificação de classes não identificáveis ao nível do *pixel* começaram a ganhar mais expressão com a cada vez maior resolução espacial das imagens adquiridas por satélites. Para fazer face às limitações da classificação ao nível do *pixel*, tem sido desenvolvida a análise de imagens orientada por objectos. A análise orientada por objectos é baseada no facto de informação semântica importante e necessária para interpretar uma imagem não ser representada em *pixels* individuais, mas em objectos significativos e nas suas relações mútuas (Gamanya *et al.*, 2007). Assim, em primeiro lugar, a imagem é segmentada em objectos. A segmentação da imagem consiste na sua subdivisão em zonas homogêneas e uniformes (objectos) em função de um determinado atributo de imagem (Woodcock *et al.*, 1994; Franklin *et al.*, 2000), dependente do método usado (e.g. *Thresholding*, segmentação

por crescimento de regiões, segmentação multi-resolução). Só depois cada um dos objectos é classificado. Para tal, a cada um é associada informação espectral, mas não só. Como cada objecto é composto por múltiplos *pixels*, podem ser usadas outras características dos objectos como a forma, tamanho, textura, relações de vizinhança, etc.

A abordagem orientada por objectos apresenta grandes vantagens em relação à abordagem ao nível do *pixel*, mas é ainda uma abordagem muito ligada à investigação e a pequenos projectos, não aos programas operacionais de cartografia de ocupação do solo. De uma forma geral, a classificação ao nível do *pixel*, e também a interpretação visual de imagens, continuam a ser os métodos mais usados em programas operacionais um pouco por todo o mundo.

#### **2.1.4. Avaliação da qualidade da cartografia**

A cartografia temática derivada de dados de detecção remota deve ser sujeita a uma avaliação rigorosa e estatística da sua exactidão. Esta avaliação pretende estimar a qualidade da cartografia para que os diferentes utilizadores deste tipo de informação possam ponderar a utilidade e viabilidade do seu uso para os fins por eles desejados (Stehman e Czaplewski, 1998). Por exactidão entende-se o grau com que a cartografia concorda com a realidade (Foody, 2002).

A avaliação da exactidão temática da cartografia começou por ser um processo simples, mas tem evoluído para maior rigor e complexidade. No início, a avaliação da exactidão era feita através de uma apreciação visual, o que é claramente subjectivo e inapropriado (Congalton, 1994 *in* Foody, 2002). Actualmente, a forma mais comum de avaliar a exactidão é baseada na comparação da cartografia com algum tipo de informação de referência, tida como correcta e verdadeira. Discordâncias entre a cartografia e a informação de referência são interpretadas como erros da primeira (Congalton, 1991 *in* Foody, 2002; Smedes, 1975 *in* Foody, 2002).

Têm sido estudados e discutidos na literatura de detecção remota muitos métodos de avaliação da exactidão (Foody, 2002), mas de uma forma geral, o processo de avaliação da exactidão de um mapa assenta na construção de uma amostra de validação através da definição dos seguintes tópicos (Congalton e Green 1999):

- i) Dados de referência;
- ii) Unidade de amostragem;
- iii) Estratégia de amostragem;
- iv) Dimensão da amostra; e
- v) Índices de avaliação da qualidade dos mapas.

A informação de referência é normalmente obtida através da interpretação visual de imagens aéreas ou vídeo (com maior resolução espacial que as imagens que deram origem

ao mapa), com base em dados de campo ou uma combinação destas fontes (Stehman e Czaplewski, 1998).

A unidade de amostragem é a unidade fundamental da avaliação e é a ligação entre a localização espacial no mapa e a correspondente localização espacial na superfície terrestre. Os dois tipos de unidade de amostragem são pontos e áreas. As áreas podem ser *pixels* ou polígonos (Stehman e Czaplewski, 1998). Normalmente é usado o *pixel* ou o polígono como unidade de amostragem, dependendo se o formato da cartografia é *raster* ou *vectorial*.

A estratégia de amostragem é o método de selecção do conjunto de unidades de amostragem (observações) para construir a amostra de validação. A adopção de uma estratégia de amostragem probabilística contribui para uma avaliação defensável do ponto de vista científico (Stehman e Czaplewski, 1998). Amostragem aleatória simples, aleatória estratificada, sistemática e por agregados, são métodos probabilísticos usados na avaliação de cartografia de ocupação do solo. Todos apresentam vantagens e desvantagens e a sua selecção deve ter em conta questões próprias da cartografia e dos dados de referência disponíveis. Por exemplo, estratégias simples, como a amostragem aleatória simples, podem ser apropriadas se a dimensão da amostra é suficiente para garantir que todas as classes de ocupação do solo são representadas adequadamente. Contudo, esta estratégia pode ser inviável se, por exemplo, não houver informação de referência disponível para toda a área cartografada (Foody, 2002).

A dimensão da amostra deve ser suficiente para que a avaliação da exactidão seja significativa (Hay, 1979 *in* Foody, 2002; Congalton e Green, 1999). De uma forma geral, quanto maior é a amostra, maior é a confiança dos resultados da avaliação da exactidão (Dicks e Lo, 1990). Contudo, a recolha de informação de referência é um processo moroso e caro, o que faz com que seja indesejável a construção de uma amostra muito grande. Assim, a amostra deve ter o tamanho mínimo que garanta o valor estatístico dos resultados (Congalton e Green, 1999). Têm sido sugeridos métodos alternativos para estimar a dimensão da amostra, a maioria baseada numa equação que aproxima a distribuição normal à distribuição binomial (Carrão e Caetano, 2006; Carrão *et al*, 2007). Contudo, Stehman (2001) demonstra que independentemente no número de *pixels* ou polígonos que compõem o mapa, uma amostra de 100 observações por classe asseguram que a exactidão pode ser estimada com um desvio padrão não superior a 0,05.

Após a selecção das observações da amostra de validação, é-lhes atribuída a *verdadeira* classe de ocupação do solo com base na informação de referência. Posteriormente, a amostra de validação é comparada com a informação temática do mapa em cada uma das observações e a magnitude com que o mapa concorda com a amostra de validação constitui a exactidão (Stehman e Czaplewski, 1998). Mais concretamente, a matriz de



confusão (também chamada de matriz de erro) é o centro da avaliação da exactidão e consiste numa tabulação cruzada entre as classes de ocupação do solo do mapa e da informação de referência. Assim, nela é descrito o padrão da distribuição de cada classe de ocupação do solo relativamente à informação de referência (Foody, 2002) através da visualização das observações na diagonal da matriz (concordância entre o mapa e a informação de referência) e fora da diagonal (confusão entre classes). Esta permite a derivação de índices que avaliam a qualidade global do mapa (e.g. exactidão global) e a qualidade da representação específica das classes de ocupação do solo (exactidão do utilizador e do produtor). A Figura 1 demonstra a construção de uma matriz de confusão de um mapa de ocupação do solo com  $q$  classes e os índices de qualidade dela derivados, em que:

$n_{ii}$  – Número de observações da amostra de validação classificadas na classe  $i$  do mapa e que corresponde à classe  $i$  na informação de referência

$n_{i+}$  – Total de observações da amostra de validação classificadas no mapa como a classe  $i$

$n_{+i}$  – Total de observações da amostra de validação pertencentes à classe  $i$

$EG$  – Exactidão Global

$EU$  – Exactidão do Utilizador

$EP$  – Exactidão do Produtor

		Informação de referência				
		1	2	...	$q$	Total
Mapa	1	$n_{11}$	$n_{12}$	...	$n_{1q}$	$n_{1+}$
	2	$n_{21}$	$n_{22}$	...	$n_{2q}$	$n_{2+}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	$q$	$n_{q1}$	$n_{q2}$	...	$n_{qq}$	$n_{q+}$
	Total	$n_{+1}$	$n_{+2}$	...	$n_{+q}$	$n$

$$EG = \frac{\sum_{k=1}^q n_{kk}}{n} \quad EU = \frac{n_{ii}}{n_{i+}} \quad EP = \frac{n_{ii}}{n_{+i}}$$

Figura 1 – Matriz de confusão com  $q$  classes de ocupação do solo e índices de qualidade do mapa (Fonte: Carrão *et al.* 2007)

## 2.2. Programas operacionais de cartografia de ocupação do solo

Existem diversos programas operacionais de cartografia de ocupação do solo por todo o mundo, com diferentes objectivos, características técnicas e área da superfície terrestre cartografada, desde pequenas a grandes escalas. No entanto, confrontados com a problemática descrita no ponto 1.1, alguns países têm vindo a desenvolver esforços para produzir cartografia de ocupação do solo à escala nacional através de métodos

automáticos. Na Tabela 1 pode-se observar um resumo das características de alguns dos principais projectos de produção de cartografia (descritos nos próximos pontos com mais pormenor) que usam classificação automática de imagens de satélite, desenvolvidos para dar resposta à desactualização da informação acerca da ocupação do solo.

	Mapa de ocupação do solo da Grã-Bretanha	CLC2000 e Mapa de Ocupação do solo Suecos	ACLUMP	NLCD 2001	Mapa de ocupação do solo do programa NR-Census
Ano de referência	1990	2000	1992-2002	2001	2004-2007
Nomenclatura	25 classes	4 e 13 classes florestais	16 classes	29 classes	19 classes
Imagens	Landsat TM	Landsat TM Landsat 7 ETM+	AVHRR	Landsat 5 Landsat 7 ETM+	AWiFS
Dados auxiliares	-	Inventário florestal nacional, COS e MDE	Censos agrícolas	MDE, imagens de satélite e informação variada	Cartografia variada
Tipo de classificador	Assistido, paramétrico	Assistido, paramétrico	Assistido, paramétrico	Assistido, não-paramétrico	Assistido, não-paramétrico
Classificador	CMV	CMV com probabilidades <i>a priori</i>	SPREADII	Árvores de decisão	Árvores de decisão
Construção da amostra de treino do classificador	Trabalho de campo	Dados auxiliares	Dados auxiliares	Dados Auxiliares e trabalho de campo	Dados Auxiliares e trabalho de campo

Tabela 1 – Projectos operacionais de cartografia de ocupação do solo que usam métodos automáticos

A referência à rapidez e/ou à contenção de custos na realização da cartografia é transversal a todos estes projectos, o que evidencia o interesse que constitui o desenvolvimento de metodologias automáticas para a produção e actualização de cartografia de ocupação do solo. O desenvolvimento que se tem observado nesta área demonstra que os métodos tradicionais, como a aplicação directa do classificador da máxima verosimilhança (CMV), podem ser superados por novos classificadores. Têm, por isso, sido substituídos por métodos mais complexos em projectos mais recentes, como os programas operacionais referidos acima revelam. Tais métodos, que incluem as árvores de decisão como classificador, estratificação pré-classificação, etc., têm sido testados no campo da investigação e comprovada as suas vantagens, e por isso, há cada vez mais possibilidades de massificar a sua aplicação em programas operacionais.

### 2.2.1. Mapa de ocupação do solo da Grã-Bretanha

No projecto de produção do mapa de ocupação do solo da Grã-Bretanha foi usada uma metodologia automática simples com o objectivo de substituir a cartografia existente, que estava bastante desactualizada, com um produto com uma unidade mínima cartográfica menor do que o projecto CORINE (Fuller *et al.*, 1994). Foi usada uma nomenclatura com 25 classes: Oceano e Águas interiores, Praia e Solo nu, várias classes de áreas artificiais, de áreas agrícolas e 18 classes de vegetação semi-natural (Tabela 2.1 do Anexo 2).

O mapa de ocupação do solo foi derivado de imagens Landsat TM, de verão e Inverno, referentes ao ano de 1990. Contudo, devido à indisponibilidade de imagens desse ano (como por exemplo, devido à existência de muita nebulosidade), foram usadas imagens com mais ou menos dois anos. A opção de usar dados multitemporais teve como objectivo ajudar a distinguir entre várias classes cujas características espectrais mudam ao longo do ano. Todavia, parte do território inglês (12%) apenas foi classificado com imagens de uma data por falta de disponibilidade de imagens. No total foram usadas 46 imagens, as quais se georeferenciaram e as imagens de verão e Inverno da mesma área foram juntas numa única composição.

Para a classificação automática das imagens, foi seleccionado o classificador da máxima verosimilhança que foi aplicado a cada composição de forma independente. Posteriormente, a classificação de todas as composições foram unidas para formar o mapa final. Os dados de treino foram recolhidos através de um intenso trabalho de campo que incluiu a recolha de informação através de transeptos estabelecidos para abranger a maior diversidade paisagística possível. Por cada imagem Landsat, foi identificada a ocupação do solo de cerca de 1200 parcelas, totalizando perto de 30000 por toda a Grã-Bretanha.

Após a classificação, o mapa produzido foi sujeito a uma edição pós-classificação baseada na criação de quatro máscaras e reclassificação de algumas ou todas as classes cobertas pelas máscaras. Em primeiro lugar, foi usada uma máscara de nuvens, as respectivas sombras e de neve. A construção desta máscara consistiu na definição de um valor limiar na banda 3 das imagens TM, acima do qual todos os *pixels* foram considerados como nuvens; na definição de um valor limiar abaixo do qual todos os *pixels* foram considerados como sombra; e na definição, mais uma vez, de um valor limiar na banda 3 das imagens TM, acima do qual todos os *pixels* foram considerados como neve. A classificação destas áreas foi substituída pela classificação de apenas uma imagem de uma data. A segunda máscara criada foi uma máscara costeira para dividir a área terrestre da área marítima e corrigir confusões entre classes nestas duas áreas (e.g. praias com áreas urbanas, oceano e águas interiores). A produção da máscara consistiu num processo semi-automático. As classes foram agregadas em terrestres e marítimas e usado um filtro da maioria numa janela de 5x5 *pixels*. As áreas resultantes deste processo com menos de 7 ha foram

removidas e outros erros que permaneceram foram editados manualmente. Após este processo, a sobreposição da máscara com a classificação permitiu corrigir confusões entre classes terrestres e marítimas. Com a terceira máscara foram separadas as terras baixas das terras altas. A separação foi realizada com base em diferenças fenológicas, topográficas e pedológicas. A eliminação de pequenas áreas isoladas de terras baixas no interior de uma grande área de terras altas e vice-versa, e a correcção de outras incorrecções, foram realizadas à semelhança da máscara costeira. A quarta máscara, uma máscara urbana/sub-urbana, permitiu eliminar os erros de comissão de algumas classes nestas áreas (nomeadamente, Terras cultivadas) devido a erros geométricos entre as imagens de Verão e Inverno. Mais uma vez, a construção da máscara foi similar à máscara costeira. As classes urbanas e sub-urbanas foram agregadas para a posterior uso de um filtro e edição manual. Por ultimo, foi ainda aplicado um filtro para eliminar os *pixels* isolados.

Não foi desenvolvido um protocolo de validação do Mapa da ocupação do solo da Grã-Bretanha. Este foi comparado com o inventário rural de 1990 (*countryside survey 1990*) incorporado na cartografia 1:10000 do *Ordnance Survey* (a agência nacional de cartografia da Grã-Bretanha). O *countryside survey 1990* é um estudo ou auditoria dos recursos naturais das áreas rurais do Reino Unido, realizado periodicamente desde 1978. Foram usados e convertidas para o formato *raster* 143 áreas de 1Km<sup>2</sup> da cartografia 1:10000 (anteriormente digitalizada) que incorporava informação do *countryside survey 1990* e definida a compatibilização entre os dados do inventário com a nomenclatura do mapa de ocupação do solo. A comparação directa entre o mapa de ocupação do solo e a informação de referência revelou uma concordância de 46%. No entanto, verificou-se que a compatibilização entre a informação de referência e o mapa de ocupação do solo era difícil de executar devido a divergentes conceitos relativos às classes de ocupação do solo. Ao usar outra configuração na compatibilização entre os dados, a concordância global foi de 67%.

### **2.2.2. CORINE Land Cover 2000 e Mapa de Ocupação do Solo suecos**

No caso do projecto CORINE Land Cover sueco, foi realizado paralelamente com o Mapa de Ocupação do Solo sueco (Hagner e Reese, 2007), nos quais foi usado um método automático para classificar as áreas florestais. A cartografia de ocupação do solo do programa CORINE tem 25 ha de unidade mínima cartográfica e usa uma nomenclatura de 44 classes no nível mais detalhado (ver ponto 2.1.1). O Mapa de Ocupação do Solo sueco é mais detalhado, tem 1 a 5 ha de unidade mínima cartográfica e usa uma nomenclatura de 57 classes, 13 das quais, florestais, de onde se originou, posteriormente, as classes florestais do CORINE Land Cover através de generalização (Tabela 2.2 do Anexo 2).

A cartografia foi derivada de imagens Landsat TM e Landsat 7 ETM+, obtidas de 1999 a 2002, e como informação auxiliar foram usados dados do inventário florestal nacional sueco, cartografia de ocupação do solo à escala 1:50000 e 1:100000 e um MDE com 50 cm de resolução espacial. Foram usadas 50 imagens de satélite, as quais foram sujeitas a pré-processamento para obter um conjunto de imagens sem nuvens e para reduzir a influência das diferenças espectrais decorrentes das diferentes datas da sua obtenção. O pré-processamento incluiu redução das diferenças de *haze* dentro de cada imagem, correcção da iluminação e modelação radiométrica.

Foi desenvolvido um sistema automatizado para reduzir ao mínimo a intervenção manual na classificação das classes florestais. O classificador usado foi uma modificação do classificador da máxima verosimilhança a que chamaram “classificação da máxima verosimilhança calibrada” (*calibrated maximum likelihood classification*). Este classificador foi desenvolvido propositadamente no âmbito do projecto e usa probabilidades *a priori*. No processo de classificação, as probabilidades *a priori* são ajustados iterativamente até alcançar a proporção esperada das classes florestais, estimadas através dos dados do inventário nacional florestal sueco. Desta forma, assegura-se que a classificação automática resulta num mapa mais exacto. O inventário nacional florestal foi também a principal fonte de informação para construir a amostra de treino, juntamente com a restante informação auxiliar.

Embora o Mapa de Ocupação do Solo sueco tenha 13 classes florestais, na fase de classificação apenas foram identificadas nove classes. Antes da classificação, foi usada a informação auxiliar para criar uma estratificação geográfica que distingue as áreas de floresta, zonas húmidas e de rocha nua, o que permitiu evitar confusões entre classes e criar a nomenclatura de 13 classes.

O Mapa de Ocupação do Solo foi avaliado com base em interpretação visual de informação de referência. Devido a limite de custos, a validação foi limitada a cinco áreas do território sueco. A amostra de validação consistiu em 3799 pontos, construída com base numa amostragem aleatória estratificada. A cartografia produzida na versão mais detalhada (Mapa de Ocupação do Solo) tem 74% de exactidão global.

### **2.2.3. Programa *Australian Collaborative Land Use Mapping***

Na Austrália foi desenvolvido um programa de cartografia do uso do solo à escala nacional (*Australian Collaborative Land Use Mapping Program*) onde as áreas agrícolas foram classificadas com imagens de baixa resolução espacial através de um método paramétrico, enquanto para os restantes usos foram usados outros dados de maior resolução.

A cartografia baseia-se na 6ª versão da nomenclatura *Australian Land Use and Management (ALUM) Classification*. Esta é uma nomenclatura detalhada e hierárquica de

três níveis, em que o primeiro e o segundo nível classificam o uso do solo e o terceiro nível dá informações úteis acerca de práticas agrícolas, vegetação, etc. No primeiro, segundo e terceiro níveis são distinguidas 6, 32 e 81 classes, respectivamente. No entanto, foi usada uma versão menos detalhada do segundo nível que consiste em 16 classes (Tabela 2.3 do Anexo 2).

A cartografia foi derivada de imagens de baixa resolução espacial (AVHRR), dados de censos agrícolas e informação acerca da ocupação do solo de fontes variadas. Este conjunto reúne um amplo recurso de informação temporal visto que está a ser preparada cartografia para os anos de 1992/93, 1993/94, 1998/99, 2000/01 e 2001/02, tendo o ano de 1996/1997 sido terminado recentemente.

O primeiro passo na produção da cartografia foi a construção de uma máscara não agrícola. Na sua construção foi usada cartografia variada como informação auxiliar: cartografia de áreas protegidas, de ocupação do solo e de floresta nativa e plantada.

A classificação baseou-se num método de modelação, o SPREADII (*Spatial REallocation of Aggregated Data*), que integra dados de censos agrícolas e dados multitemporais de imagens de satélite para classificar as áreas agrícolas. O princípio base seguido foi que o uso do solo está associado à ocupação do solo, que pode ser estimada através de perfis espectrais. O SPREADII usou séries temporais do *Normalized Difference Vegetation Index* (NDVI) de imagens AVHRR para construir os perfis espectrais através de áreas georeferenciadas constantes numa base de dados de treino. Esta base de dados foi fornecida pelos dados censitários cujos usos do solo eram conhecidos. Os perfis espectrais foram construídos por um conjunto de parâmetros: altura do ano com NDVI máximo, NDVI médio e a diferença entre os NDVI máximo e mínimo. Para proceder à classificação, para cada *pixel*, é calculada a probabilidade de apresentar um dado uso do solo mediante a comparação dos seus parâmetros com os parâmetros de todas as áreas de treino que correspondem ao uso do solo a comparar. O SPREADII atribui repetidamente um uso do solo a cada *pixel* para um determinado número de iterações. No fim do processo, o número de vezes que um a *pixel* foi atribuído um determinado uso do solo, dividido pelo número de iterações, resulta na probabilidade estimada que esse *pixel* apresente esse uso do solo (BRS, 2006).

#### **2.2.4. Projecto *National Land-Cover Database* (NLCD 2001)**

Abordagens não convencionais, nomeadamente o uso de árvores de decisão, foram já adoptadas em programas operacionais, apesar do seu recente uso na área da detecção remota. Por exemplo, no âmbito do projecto norte-americano *National Land-Cover Database project* de 2001 (NLCD 2001) foi derivado automaticamente um mapa de ocupação do solo de 29 classes (Tabela 2.4 do Anexo 2) com árvores de decisão e imagens Landsat 5 e Landsat 7 ETM+ (Homer *et al.*, 2004).

A área foi submetida a uma estratificação geográfica pré-classificação baseada nas ecoregiões delimitadas por Omernik (1987), na cartografia NLCD 1992 e em dados AVHRR de onde resultaram 66 estratos nos estados continentais e 23 estratos no Alaska.

O projecto foi desenvolvido com imagens de três alturas do ano de 2001 (antes, durante e depois da época em que a vegetação tem maior vigor) para se efectuar uma avaliação multi-temporal como um indicador da fenologia da vegetação. Devido à grande quantidade de imagens necessárias para abranger toda a área, depois de georeferenciadas, a correcção radiométrica das imagens Landsat 7 foi realizada através dos métodos *standart* do *U.S. Geological Survey*. As imagens Landsat 5 foram primeiramente convertidas radiometricamente para imagens Landsat 7 através dos coeficientes desenvolvidos por Vogelmann *et al.* (2001). As imagens foram convertidas para a reflectância no topo da atmosfera para todas as bandas reflectivas, para a temperatura no topo da atmosfera para as bandas térmicas e submetidas à transformação *tasseled cap*. Por último, para cada estrato, construiu-se um mosaico com todas as imagens.

Como informação auxiliar usada na classificação foi usado um MDE e outra informação auxiliar adicional para criar uma máscara urbana (densidade populacional, *buffers* da rede viária, o NLCD 1992 e áreas iluminadas à noite através de imagens do *National Oceanic and Atmospheric Administration* (NOAA)).

À parte da cartografia de ocupação do solo, foram criados outros dois produtos, o grau de impermeabilização do solo e a densidade da cobertura vegetal. A informação relativa à impermeabilização do solo foi escolhida para classificar a densidade urbana no mapa final. Foram usados quadrados de imagens com 1 m de resolução espacial para criar dados de referência de áreas impermeáveis para depois calibrar a relação entre estas e as imagens Landsat. O modelo foi depois aplicado a toda a área e produzido um mapa com a estimativa da percentagem de área impermeável ao nível do *pixel*. A mesma metodologia foi aplicada à densidade de cobertura vegetal.

Para classificar as imagens foi usada uma árvore de decisão. De uma forma geral, os dados de treino foram obtidos de programas estatais e federais existentes e complementados por trabalho de campo em áreas sem dados disponíveis. Foram usadas ortoimagens de alta resolução espacial, bases de dados locais, trabalho de campo e do *Forest Inventory Analysis*. Em alguns estratos, foram ainda usados dados de cartografia regional de ocupação do solo (e.g. *Gap Analysis Program* e dados agrícolas do *National Agricultural Statistics Service*).

Após a classificação, a cartografia foi editada para alterar a unidade mínima cartográfica para 1 acre, através de agregação. Este processo foi realizado com um algoritmo que agregou as áreas com menos de 1 acre à área vizinha mais semelhante, em função de uma

tabela de prioridades de agregação. Por ultimo, todos os estratos foram unidos para formar a cartografia final.

Os resultados da validação deste produto ainda não são conhecidos. Contudo, foi realizada uma validação cruzada referente a 65 estados (que não incluem o Alasca, Havai e Porto Rico) de forma a ter uma ideia da exactidão da cartografia produzida. O processo foi realizado através do algoritmo usado na classificação, em que os dados de treino foram divididos em 10 subconjuntos de igual tamanho. Cada um dos conjuntos foi classificado com base dos restantes nove, obtendo-se uma estimativa da exactidão da classificação. Estima-se que a classificação tenha uma exactidão de 70% a 98%, dependendo do estrato, com uma exactidão média de 83.9%.

### **2.2.5. Mapa de ocupação do solo do programa indiano NR-Census**

Desde 2004, o Departamento do Espaço do governo indiano, no âmbito do programa *Natural Resource Census* (NR-Census) inserido no *National Natural Resources Management System*, tem desenvolvido e implementado métodos automáticos para produzir anualmente cartografia de ocupação do solo à escala 1:250000 (NRSA, 2005). No projecto definiram-se 19 classes de ocupação do solo (Tabela 2.5 do Anexo 2) para derivar cartografia da Índia a partir de imagens AWiFS.

Até ao momento foi produzida cartografia para os anos de 2004/2005, 2005/2006 e 2006/2007. Para cada data, foram usadas, respectivamente, 680, 762 e 802 quadrantes de imagens AWiFS adquiridas durante 10 meses para cada época de produção agrícola na Índia: *kharif* (Agosto-Outubro), *Rabi* (Janeiro-Março) e *Zaid* (Abril-Maio). As imagens foram georeferenciadas e convertidas para a reflectância no topo da atmosfera para minimizar a variabilidade temporal. Para cada estado indiano, construiu-se um mosaico das imagens necessárias para cobrir toda a área e, assim, fazer a cartografia estado a estado. Nas áreas cobertas de nuvens, substituíram-se as imagens AWiFS por imagens WiFS e MODIS.

Como informação auxiliar foram usadas cartografia topográfica e a cartografia temática acerca das florestas e zonas húmidas.

Para a classificação automática dos dados multitemporais AWiFS, seleccionou-se as árvores de decisão. Na fase de classificação, a informação auxiliar serviu para construir máscaras de floresta, zonas húmidas, etc. Adquiriu-se dados de campo para usar na fase de treino através de uma amostragem aleatória estratificada. A estratificação baseou-se em mapas topográficos, NDVI, numa classificação não assistida, e outra informação.

Na validação usou-se igualmente uma amostragem aleatória estratificada para construir a amostra de validação e usou-se a informação auxiliar e dados de campo como informação de referência. A cartografia produzida para os três momentos temporais alcançou elevadas percentagens de exactidão global. A cartografia de 2004/2005 apresenta 83% a 87% de



exactidão global, dependendo do estado. A cartografia de 2005/2006 atingiu 86% a 95% e a cartografia de 2006/2007 atingiu 83% a 95%.

### 3. Metodologia

De forma a atingir os objectivos do projecto COSMIC, foi desenhada uma aproximação metodológica simples e convencional de classificação automática de um conjunto intra-anual de imagens de satélite AWiFS (Tabela 2).

Resolução espacial (m)	Resolução espectral ( $\mu\text{m}$ )	Resolução radiométrica (bits)	Swath (Km)
60	0.52 – 0.59 (Verde)	10	370
	0.62 – 0.68 (Vermelho)		
	0.77 – 0.86 (Infra-Vermelho)		
	1.55 – 1.70 (IV Médio)		

Tabela 2 – Características das imagens AWiFS (Fonte: NRSA, 2005)

Para obter informação acerca do todo o território continental, são necessários dois quadrantes<sup>1</sup> de imagens AWiFS (Norte e Sul), que foram adquiridas para os meses de Abril, Julho e Outubro de 2006, totalizando seis imagens (Figura 2).

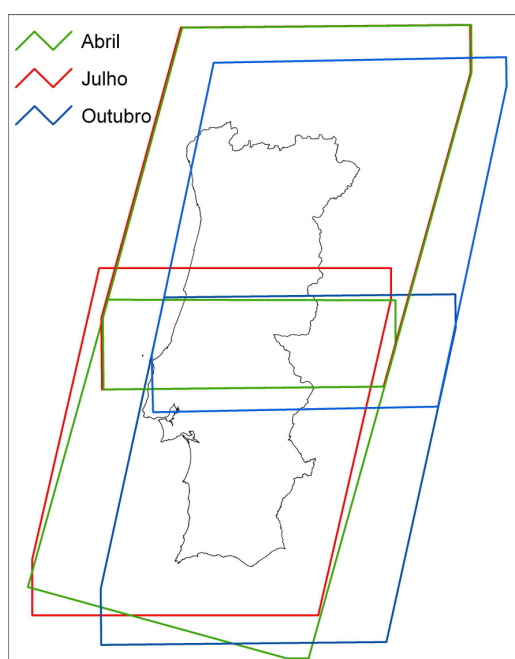


Figura 2 – Área coberta pelas seis imagens AWiFS

Com base nas imagens disponíveis, a definição da aproximação metodológica teve em conta a sua estrutura simples, muito usada na área da cartografia de ocupação do solo derivada de imagens de satélite, cuja eficiência tem sido demonstrada através de vários projectos análogos executados com estruturas similares. Dentro da aproximação

---

<sup>1</sup> As imagens do sensor do AWiFS são divididas em quatro partes e distribuídas por quadrantes

metodológica, desenhada em cinco fases, foi pensada a definição de diferentes abordagens de classificação de forma a testar várias opções metodológicas que melhor se ajustassem aos objectivos do projecto COSMIC. Todo o processo foi desenvolvido de forma interactiva entre as fases constituintes, de forma a recuar no processo com base nos resultados obtidos e redefinir e testar diferentes abordagens e opções metodológicas. A metodologia é constituída pelas seguintes fases:

- i) Estratificação geográfica;
- ii) Pré-processamento das imagens;
- iii) Definição da nomenclatura de ocupação do solo;
- iv) Classificação assistida;
- v) Avaliação da exactidão do mapa.

A metodologia começou pela elaboração de uma estratificação pré-classificação de Portugal Continental com base em trabalhos de caracterização da paisagem portuguesa. Esta fase teve como objectivo isolar características espectrais da superfície terrestre e do respectivo padrão de ocupação do solo que colocam problemas a uma classificação automática. Ao estratificar o território, é possível tratar cada estrato individualmente e melhorar a exactidão da classificação.

A segunda fase consistiu no pré-processamento das imagens AWiFS. Foram corrigidas distorções geométricas através da ortorectificação das imagens e identificadas áreas cobertas por nuvens e sombra. Devido a estas áreas estarem impossibilitadas de fornecer informação relativa às três datas das imagens, é obrigatória a sua identificação e exclusão e aplicar um procedimento alternativo na sua classificação.

Foi adoptada uma estratégia multi-temporal de classificação assistida de três imagens de satélite AWiFS. Foram usadas imagens de três alturas do ano, de Primavera, Verão e Outono (Tabela 3), distintas em termos de fenologia da vegetação e práticas do uso do solo, para aumentar a informação sobre os diferentes estados espectrais da paisagem, e assim, melhor distinguir as classes de ocupação do solo (Figura 3).

Imagem	Data de aquisição	Elevação solar	Azimute solar
Primavera (Norte)	27/04/2006	60.6 <sup>0</sup>	154.2 <sup>0</sup>
Verão (Norte)	8/07/2006	67.7 <sup>0</sup>	142.7 <sup>0</sup>
Outono (Norte)	13/10/2006	39.7 <sup>0</sup>	161.8 <sup>0</sup>
Primavera (Sul)	27/04/2006	63.0 <sup>0</sup>	149.6 <sup>0</sup>
Verão (Sul)	28/07/2006	64.0 <sup>0</sup>	133.0 <sup>0</sup>
Outono (Sul)	13/10/2006	42.1 <sup>0</sup>	160.3 <sup>0</sup>

Tabela 3 – Parâmetros de aquisição das imagens AWiFS

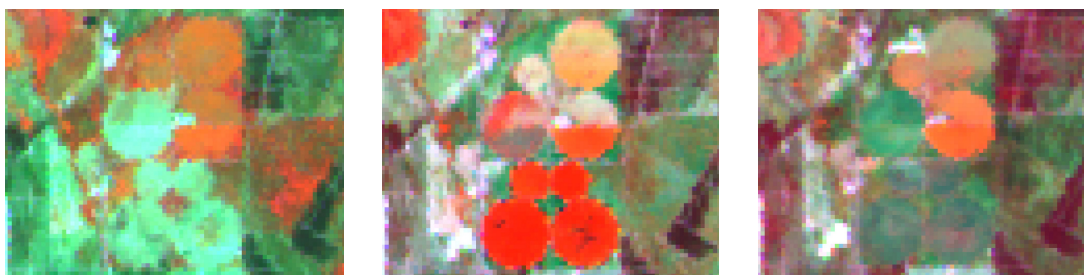


Figura 3 – Exemplo de diferentes estados espectrais ao longo do ano de áreas agrícolas irrigadas

Foi também usada diversa informação auxiliar nas fases de pré-processamento das imagens, classificação assistida e avaliação da qualidade dos resultados (Tabela 4).

Informação auxiliar	Data de referência	Proprietário	Escala	Resolução espacial /unidade mínima cartográfica	Formato
MDE	-	IGP	1:100 000	50 m	<i>Raster</i>
Mosaico de imagens LISS III e SPOT	2006	IGP		20 m	<i>Raster</i>
Ortoimagens	2004 a 2006	IGP/DGRF	1:5 000	0.5 m	<i>Raster</i>
Inventário de vinhas	2003	IVV		-	Vectorial
COS'90	1990	IGP	1:25 000	1 ha	Vectorial
CLC2000	2000	IA	1:100 000	25 ha	Vectorial

Tabela 4 – Informação auxiliar

Na terceira fase, foi definida a nomenclatura de ocupação do solo através de um processo interactivo ao longo do trabalho de forma a usar a nomenclatura mais detalhada possível em função das capacidades discriminativas das imagens AWiFS.

Numa classificação automática de imagens, o classificador tem um papel fundamental. De entre os vários classificadores usados em detecção remota, foi seleccionado um classificador paramétrico, o classificador da máxima verosimilhança, e um classificador não paramétrico, árvores de decisão. A selecção de dois classificadores está relacionada com a exploração das imagens AWiFS para avaliar a sua capacidade em classificar a paisagem portuguesa. Deste modo, a selecção de dois classificadores permite corroborar conclusões e despistar se as mesmas são unicamente originadas por questões metodológicas. Por outro lado, vai ao encontro dos recentes desenvolvimentos nesta área de aplicação da detecção remota, onde se têm obtido melhores resultados com classificadores não paramétricos, entre os quais, as áreas de decisão.

Foi ainda seleccionado o método de classificação ao nível do *pixel* como o principal método de classificação com uma unidade mínima cartográfica de 3600 m<sup>2</sup>, que corresponde à resolução espacial das imagens AWiFS. Esta selecção deveu-se ao facto da classificação

ao nível do *pixel* ser o método de classificação automático mais simples, o que permite dois aspectos: por um lado, celeridade na execução da metodologia, por outro lado, ao não incorporar complexas considerações metodológicas, realça as conclusões da avaliação das imagens AWiFS quanto à sua capacidade em classificar a paisagem portuguesa. Além disto, esta opção não inviabiliza a evolução da metodologia para uma classificação ao nível do *pixel* combinada com informação contextual, a generalização da cartografia para uma maior unidade mínima cartográfica e a sua conversão para o formato vectorial, muitas vezes mais aceite por parte do utilizador. Desta forma, foi ainda testada uma classificação híbrida que combina a classificação ao nível do *pixel* e análise orientada por objectos.

No fim, a exactidão dos resultados produzidos ao longo do trabalho foram avaliados através de uma avaliação probabilística. Na Figura 4 pode-se observar uma representação geral de toda a metodologia e nos pontos seguintes é apresentada em pormenor cada uma das suas fases.

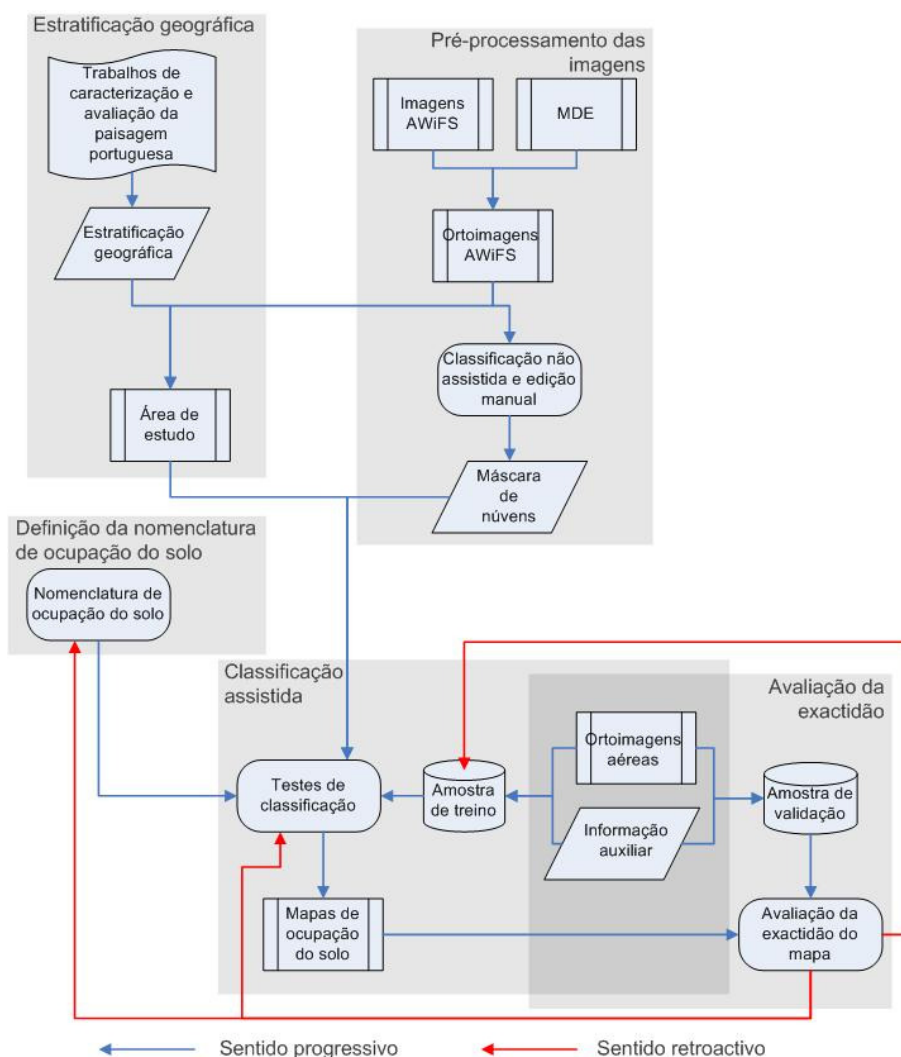


Figura 4 – Representação geral da metodologia

### 3.1. Estratificação geográfica

A metodologia teve em consideração problemas comuns em classificação automática de imagens de satélite que se prendem com diferentes características espectrais da superfície terrestre e do respectivo padrão de ocupação do solo, em função do espaço geográfico e do número de imagens que cobrem a área a classificar. O uso de diversas imagens que abrangem grandes áreas proporciona uma variedade de padrões espectrais intra e entre imagens relacionados com a fisiografia, fenologia, condições de iluminação, atmosféricas e do próprio sensor (Homer *et al.*, 2004).

No caso de Portugal Continental, não é necessário um grande número de imagens AWiFS para cobrir todo o território. No entanto, é reconhecida a grande diversidade paisagística que o caracteriza, pelo que continua a existir grande diversidade espectral. Desta forma, torna-se bastante pertinente diferenciar a área de estudo em função das suas características paisagísticas. Isto faz com que seja possível compartimentar padrões de características e problemas espectrais regionais, o que permite implementar procedimentos distintos e personalizados adaptados a tais características e problemas, que por sua vez possibilita a obtenção de resultados com maior qualidade. Com este objectivo, é comum criar uma estratificação com base nas características espectrais e em critérios biofísicos para diferenciar as diversas características do território (Homer *et al.*, 2004), o que pode aumentar a exactidão temática da cartografia entre 10% a 15% (Bauer *et al.*, 1994). Assim, foi realizada uma estratificação do território continental, o que implicou delimitar regiões homogéneas para adaptar a estratégia metodológica a cada uma das regiões em função das respectivas características com o objectivo final de maximizar a exactidão temática da carta de ocupação do solo a produzir.

O desenho de regiões homogéneas não é um processo simples. Implica objectivos em relação ao seu futuro uso, escala de análise, informação de base, capacidades analíticas, etc. Todos estes factores fazem com que o processo seja muito exigente em recursos humanos, de conhecimento, tempo, e custos económicos. Assim, faz sentido partir de trabalhos previamente realizados em que de alguma forma seja feita uma delimitação de regiões homogéneas para Portugal.

De entre diversos trabalhos de caracterização do território português e sua delimitação em regiões homogéneas, o trabalho *Contributos para a Identificação e Caracterização da Paisagem em Portugal Continental* (DGOTDU, 2004) apresentou-se como uma base de grande potencialidade por ser transversal e completo no que diz respeito aos critérios utilizados e por ser relativamente recente. Neste trabalho, Portugal é dividido em 121 unidades de paisagem (Figura 5) e teve em consideração variáveis como a geomorfologia, litologia, solos, uso do solo, dimensão das explorações agrícolas e padrão do povoamento, que são factores bastante relevantes na classificação de imagens.

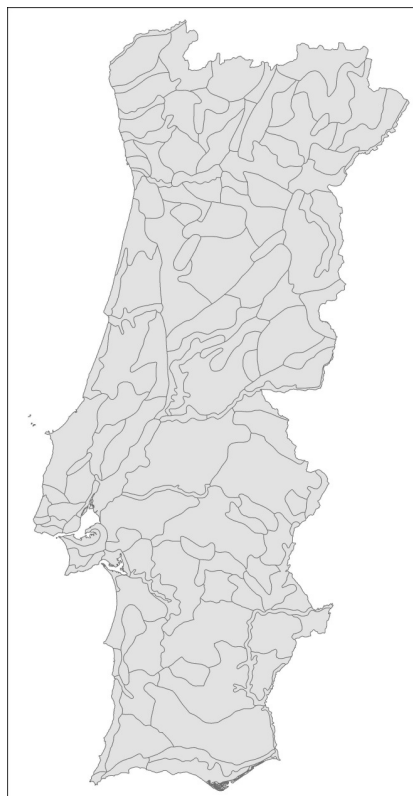


Figura 5 – Unidades de paisagem (Fonte: DGOTDU, 2004)

Com efeito, a caracterização e divisão do território nas unidades de paisagem apresentada no referido trabalho traduzem já as características e problemas inerentes ao território e aos seus padrões espectrais. Estes argumentos tornam o trabalho citado numa base proveitosa para definir uma estratificação de Portugal no âmbito de uma classificação assistida. Assim, as unidades de paisagem apresentadas foram reinterpretadas à luz dos objectivos do projecto COSMIC e definida uma estratificação para Portugal Continental.

A 121 unidades de paisagem foram reinterpretadas e agregadas porque o seu uso directo como apresentadas no trabalho original seria inviável. A divisão territorial em regiões homogéneas possibilita o aumento da exactidão da classificação, mas também aumenta os custos de produção. Logo, a estratificação teve que ser realizada de forma a obter um compromisso entre o detalhe da estratificação e o aumento dos custos. Uma sobre-estratificação poderá levar à inviabilidade da execução do trabalho em tempo útil e/ou a custos incompressíveis, contrariando o enquadramento geral da metodologia. Por este motivo, a estratificação geográfica foi realizada em três níveis hierárquicos. No primeiro nível, Portugal Continental foi dividido em apenas dois estratos, que separam as áreas com maior diversidade de ocupação do solo do Sul das áreas mais monótonas e agrestes do Norte. No segundo nível, estes dois estratos foram divididos em mais dois estratos que separam as áreas litorais das áreas interiores. No terceiro nível, o território foi estratificado num total de 13 estratos (Figura 6 e Tabela 5).

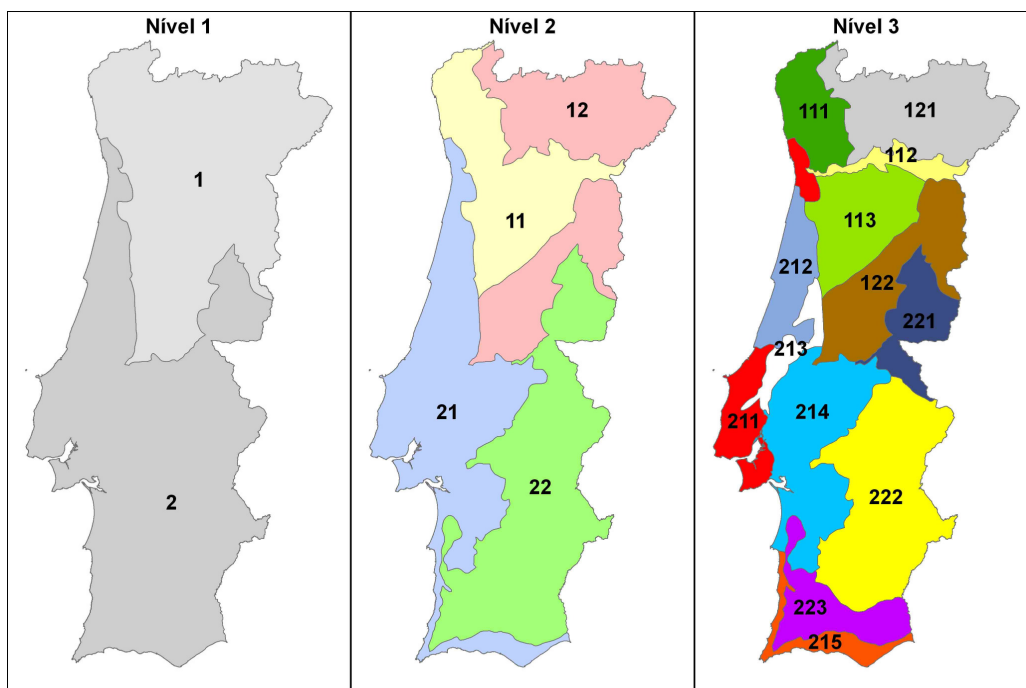


Figura 6 – Estratificação de Portugal Continental

Nível 1	Nível 2	Nível 3	Características gerais
1 Terras altas do Norte	11 Litoral verdejante do Norte	111	Paisagem sempre verde com forte presença humana
		112	Vale do Douro com diversidade de usos, maioritariamente agrícola
		113	Pequenas serras com floresta e alguma agricultura diversificada
	12 Interior agreste do Norte	121	Terrenos desertificados e agrestes (afloramentos rochosos, floresta e matos em áreas de montanha) e áreas agrícolas nas zonas baixas
		122	Terrenos rudes e extensivos, com florestas monótonas, afloramentos rochosos e pouca vegetação. Pouca diversidade
2 Terras baixas do Sul	21 Litoral diversificado do Sul	211	Grandes áreas urbanas e suburbanas, áreas naturais, grande diversidade de uso, paisagens incaracterísticas
		212	Plano arenoso com uso florestal e agricultura diversificada
		213	Serras calcárias com usos diversificados
		214	Áreas planas com diversidade de usos e grande presença do Tejo e Sado
		215	Áreas costeiras de praia e arribas, com uso diversificado do solo
	22 Interior extensivo do Sul	221	Transição entre Alentejo e Beira
		222	Vasta planície com alguns acidentes e usos agro-silvo-pastoris
		223	Áreas de serra monótona com domínio de matos e matas

Tabela 5 – Níveis de estratificação geográfica e suas características gerais



Após a estratificação, um estrato foi seleccionado para ser usado como área de estudo. O estrato seleccionado foi o 214 (Figura 6) devido às suas características, como a variedade de tipos de ocupação do solo e a heterogeneidade da paisagem. A sua área é cerca de 1223890 ha e cobre grande parte dos vales dos rios Tejo e Sado. O objectivo desta selecção foi garantir que os problemas que surgissem durante a classificação da área de estudo fossem demonstrativos dos problemas colocados pelos restantes estratos. O trabalho e resultados descritos na presente dissertação estão relacionados somente com a área de estudo (estrato 214).

### 3.2. Pré-processamento das imagens

Antes de se avançar no trabalho, foi necessário proceder à ortorectificação das imagens AWiFS para reduzir a distorção topográfica provocada pelo relevo. As imagens encontravam-se georeferenciadas ao datum WGS84 e projectadas na projecção Cónica Conforme de Lambert. Foram ortorectificadas para o sistema PT-TM06-ETRS89 com recurso ao MDE produzido pelo IGP com 50 m de resolução espacial (Tabela 6).

	Imagens originais	Imagens ortorectificadas
Elipsoide	WGS84	GRS80
Datum	WGS84	ETRS89
Projecção	Cónica Conforme de Lambert	Transversa de Mercator

Tabela 6 – Sistemas de representação das imagens antes e após a ortorectificação

O processo de ortorectificação, realizado com a aplicação comercial PCI Geomatics®, consistiu na recolha de pontos de controlo, cujas coordenadas são conhecidas, com base num mosaico de imagens LISS-III e SPOT 4, previamente ortorectificadas para o sistema PT-TM06-ETRS89. Depois de recolhidos os pontos de controlo, as imagens foram ortorectificadas por funções racionais com um erro inferior a meio *pixel* (Tabela 7 e Anexo 1).

	Abril Sul	Julho Sul	Outubro Sul	Abril Norte	Julho Norte	Outubro Norte
RMS	0.39	0.42	0.42	0.42	0.39	0.43
XMRS	0.28	0.28	0.31	0.29	0.26	0.30
YMRS	0.28	0.31	0.28	0.31	0.29	0.32

Tabela 7 – Erro médio quadrático (RMS) da ortorectificação das imagens AWiFS

Foram identificadas áreas cobertas por nuvens e as respectivas sombras em cada uma das imagens. Para o efeito, cada imagem foi sujeita a uma classificação não assistida com o classificador K-means. Os *clusters* correspondentes às nuvens e sombras foram identificados, isolados e corrigidos, quando necessário, por edição manual. A área que resultou foi ainda sujeita a uma operação de alargamento (*buffer*) de um *pixel* para

assegurar que toda a área coberta por nuvens e sombra era seleccionada. No fim, as áreas identificadas como nuvens e sombra de cada imagem foram agrupadas num único tema de informação. Deste procedimento resultou uma máscara de exclusão.

Curiosamente, verificou-se que no estrato 214 não existem nuvens em qualquer uma das imagens. De qualquer forma, futuramente esta máscara tem como função evitar amostrar nas áreas afectadas por nuvens na a fase de treino do classificador, aquando da classificação de todo o país. Por este motivo, a classificação das áreas incluídas na máscara de exclusão seguirá uma via alternativa e personalizada que consistirá em usar menos informação temporal. Por outras palavras, nestes casos, a classificação usará apenas as imagens que não apresentam nuvens. Por exemplo, uma área que esteja incluída na máscara de exclusão por apresentar nuvens ou sombra na imagem de Abril, apenas poderá ser classificada com a informação espectral das imagens de Julho e Outubro.

### **3.3. Nomenclatura de ocupação do solo**

A nomenclatura de ocupação do solo não foi definida pela aplicação directa (sem alterações) de nenhum sistema de classificação de ocupação do solo pré-existente. A nomenclatura LANDEO foi seleccionada como ponto de partida para construir uma nomenclatura adaptada. Como referido no ponto 2.1.1, a nomenclatura LANDEO é uma nomenclatura hierárquica, detalhada e multiescala, desenvolvida pelo GDR para Portugal Continental com o propósito de ser usada em cartografia de ocupação do solo derivada de métodos automáticos (Araújo e Caetano, 2006). Assim, a nomenclatura original (Tabela 4.1 do Anexo 4) foi adaptada e simplificada ao longo do trabalho num processo interactivo, de acordo com a resolução espacial e as capacidades das imagens em discriminar a paisagem portuguesa por via de testes preliminares. Neste processo interactivo, os pomares e olivais foram excluídas devido à confusão com outras classes. Verificou-se que a estrutura aberta destas áreas agrícolas faz com que o sinal espectral registado pelo sensor seja muito influenciado pelo sobcoberto, o que provoca confusões com classes que apresentam a mesma ocupação ao nível do solo (e.g. Vegetação herbácea natural). Por outro lado, a classe Mosaicos, existente na nomenclatura LANDEO, não fazia sentido em função da resolução espacial das imagens (60 m), pelo que não foi contemplada. Deste processo resultou uma nomenclatura de 15 classes de ocupação do solo (Tabela 8). Esta nomenclatura pode parecer pouco detalhada relativamente a outras nomenclaturas, nomeadamente às nomenclaturas usadas nos programas descritos no ponto 2.2. Todavia, à que notar que algumas classes usadas nos referidos programas não têm muito interesse para o projecto COSMIC (e.g. distinguir águas costeiras de águas interiores) e outras classes não são muito relevantes no território português (por exemplo, em Portugal não existe uma forte dicotomia entre floresta de folhosas caducas e folhosas persistentes, como

noutros países). Além disso, nos programas referidos, a distinção de muitas classes não é feita através da classificação das imagens de satélite, mas sim com base em máscaras (e.g. águas costeiras e águas interiores). No presente trabalho, também se poderia adoptar uma estratégia semelhante, mas o maior detalhe obtido na nomenclatura não estaria relacionado com uma maior eficiência metodológica ou com a informação oferecida pelas imagens AWiFS.

Código	Classe de ocupação do solo
1	Áreas artificiais contínuas
2	Áreas artificiais descontínuas
3	Culturas de sequeiro
4	Culturas de regadio
5	Arrozais
6	Vinhas
7	Florestas de folhosas
8	Florestas de resinosas
9	Áreas agro-florestais
10	Matos
11	Vegetação herbácea natural
12	Cortes florestais
13	Zonas húmidas
14	Solo nu
15	Água

Tabela 8 – Nomenclatura de ocupação do solo

No Tabela 4.2 do Anexo 4 pode-se consultar a correspondência entre as 15 classes da Tabela 8 e a nomenclatura LANDEO original e ainda a respectiva descrição.

A nomenclatura da Tabela 8 foi definida como a nomenclatura padrão, mas para além dela foram testadas outras versões menos detalhadas, as quais são referidas oportunamente.

### 3.4. Classificação assistida

A classificação das imagens foi a fase nuclear do trabalho desenvolvido porque incluiu várias tarefas e a definição das várias abordagens de classificação. Influenciou a definição da nomenclatura de ocupação do solo num processo interactivo, como referido anteriormente; incluiu a selecção dos classificadores; a definição da amostragem de treino; e a definição e execução de diversos testes de classificação com diferentes variáveis (classificador, classes de treino, etc.) para avaliar a melhor abordagem metodológica. As características dos dois classificadores seleccionados, a amostragem de treino e os testes de classificação são apresentados em pormenor nos próximos três pontos.

### 3.4.1. Os classificadores da máxima verosimilhança e árvores de decisão

Os classificadores assistidos mais comuns em *software* de detecção remota são os classificadores paramétricos, tais como os classificadores da mínima distância, do paralelepípedo e da máxima verosimilhança, o que os tornam bastante acessíveis e facilmente aplicados. Além disso, a sua simplicidade, fundamentação estatística e a longa experiência na sua aplicação, tornam estes classificadores ainda hoje usados em classificação de dados de detecção remota. De entre este grupo, o classificador da máxima verosimilhança é definitivamente o mais usado (e.g. Fuller *et al.*, 1994; El-Magd e Tanton, 2003; Dean e Smith, 2003; Hagner e Reese, 2007; Laba *et al.*, 2008, Shalaby e Tateishi, 2007) e um dos mais exactos. No entanto, os classificadores paramétricos têm sido criticados em alguns pontos: i) fazem pressupostos acerca da distribuição dos dados, na sua maioria, a distribuição normal (Pal e Mather, 2003), o que pode não ser correcto em imagens de detecção remota que contêm *pixels* mistos (Xu *et al.*, 2005); ii) as características dos dados de treino exercem uma influência considerável na exactidão da classificação (Campbell, 1981); iii) o tamanho da amostra de treino tem que ser grande e aumenta rapidamente juntamente com o aumento das variáveis de *input* (do inglês *input features*) para evitar a maldição da dimensionalidade dos dados, também conhecido como *fenómeno de Hughes* (Hughes, 1968). Apesar das suas limitações, os classificadores paramétricos continuam a ser muito usados.

O classificador da máxima verosimilhança é um classificador que calcula a probabilidade dos indivíduos a classificar pertencerem às classes presentes na fase de treino. Cada indivíduo é classificado na classe na qual apresenta maior probabilidade de pertencer, onde a probabilidade  $L_i(x)$  de um indivíduo  $x$  ser membro da classe  $i$  é dado por (Dean e Smith, 2003):

$$L_i(x) = (2\pi)^{-n/2} |V_i|^{-1/2} e^{-y/2}$$

onde  $V_i$  é a matriz de covariância da classe  $i$ ,  $n$  é o número de bandas espectrais e  $y$  é a distância de Mahalanobis.

O classificador da máxima verosimilhança assume que os dados de treino são normalmente distribuídos (distribuição de Gauss) em cada banda das imagens (Blaisdell, 1993 *in* Jensen, 1996). Assim, dados de treino com histogramas bi ou tri-modais numa banda não são adequados. Nestes casos, cada uma das modas pode representar diferentes classes que devem ser treinadas separadamente (Jesen, 1996).

Recentemente, têm sido desenvolvidos outros classificadores de forma a ultrapassar alguns problemas colocados pelas abordagens paramétricas. Os classificadores não paramétricos

não fazem pressupostos acerca da distribuição dos dados e têm sido descritos como mais exactos (Friedl e Brodley, 1997; Xu *et al.*, 2005; Pal e Mather, 2003; Rogan *et al.*, 2002). As árvores de decisão são um destes classificadores e têm sido descritos como vantajosos devido a várias características, nomeadamente: a estrutura de classificação relativamente simples, explícita e intuitiva (Friedl e Brodley, 1997); a capacidade de lidar com relações não lineares entre as variáveis de *input* (Friedl e Brodley, 1997; Xu *et al.*, 2005), com dados representados em diferentes escalas de medida (Pal e Mather, 2003) e com dados contínuos e discretos de forma simultânea (Homer *et al.*, 2004); maior rapidez na fase de treino (Gahegan e West 1998 *in* Pal e Mather, 2003; Homer *et al.*, 2004) e no cálculo computacional (Pal e Mather, 2003; Homer *et al.*, 2004).

Por outro lado, a influência do tamanho da amostragem de treino tem sido descrita de forma inconsistente. Pal e Mather (2003) afirmam que estes classificadores não requerem uma amostra de treino muito grande para serem eficazes enquanto Joy *et al.* (2003) *in* Sesnie *et al.* (2008) afirmam o contrário. Oates e Jenson (1997) *in* Pal e Mather (2003) sugerem que o tamanho da amostra de treino não influencia a exactidão da classificação. Também de forma inconsistente tem sido descrita a resposta a *outliers*. Friedl e Brodley (1997) afirmam que as árvores de decisão são flexíveis e robustas relativamente a relações ruidosas entre as variáveis de *input* e as classes, enquanto Miller e Franklin (2002) *in* Sesnie *et al.* (2008) afirmam que a instabilidade das árvores de decisão aos *outliers*, ou a pequenas mudanças nos dados de treino, são uma limitação.

As vantagens das árvores de decisão estão relacionadas com a sua estrutura. Ao contrário dos classificadores comuns que usam os dados de treino para classificar num única etapa, as árvores de decisão são baseadas numa estrutura multi-etápica e hierárquica ou estrutura em árvore. A árvore é composta por um nó raiz (que contém todos os dados), um conjunto de nós internos e um conjunto de nós terminais (*folhas*). Cada nó tem apenas um nó ascendente e dois ou mais nós descendentes. Em cada nó da estrutura há uma separação do conjunto de dados em subconjuntos mais pequenos e homogéneos com base num conjunto de testes. Existem vários algoritmos para realizar os testes, os mais conhecidos são o CART (*Classification and Regression Trees*), CHAID (*Chi Square Automatic Interaction Detection*), C4.5/C4 e o ID3. O processo ocorre em todos os nós ao longo da estrutura até alcançar um nó terminar, o qual está associado a uma classe (Xu *et al.*, 2005; Friedl e Brodley, 1997).

Outro aspecto que favorece as árvores de decisão é a possibilidade de usar as técnicas *pruning* e *boosting*. *Pruning* é numa técnica para evitar que a árvore de decisão se ajuste demasiado à amostra de treino, o que reduz o seu desempenho em contacto com os futuros dados a classificar e consiste na redução do tamanho da árvore, eliminando ramos terminais. *Boosting* é uma técnica para melhorar a exactidão da árvore de decisão. Após a construção de uma árvore, é possível verificar que observações de treino foram mal

classificadas. Estas observações são usadas a seguir com maior peso da construção de outra árvore, o que faz com que esta dê mais atenção a estes casos e os classifique correctamente, e assim sucessivamente. No final, os dados são classificados pela ponderação do resultado de várias árvores de decisão, produzindo melhores resultados do que o uso de apenas uma árvore de decisão.

A tendência para usar outros classificadores, como as árvores de decisão, em vez do classificador da máxima verosimilhança, deve-se aos melhores resultados obtidos com os primeiros. Vários trabalhos têm sido publicados em que se compara os dois classificadores na produção de cartografia de ocupação do solo. Friedl e Brodley (1997) testaram três algoritmos de árvores de decisão em três conjuntos de imagens de detecção remota e, em todos os casos, obtiveram melhores resultados comparativamente ao classificador da máxima verosimilhança. Rogan *et al.* (2002) testaram os dois classificadores na monitorização e levantamento cartográfico das alterações da cobertura vegetal no sul da Califórnia entre 1990 e 1996. As árvores de decisão permitiram obter resultados superiores em cerca de 10% na exactidão da classificação. Xu *et al.* (2005) obtiveram igualmente melhores resultados na classificação automática de imagens Landsat ETM+ com árvores de decisão, comparativamente ao classificador da máxima verosimilhança, ambos aplicados como classificadores relaxados. Por outro lado, uma das desvantagens das árvores de decisão é a sua ausência generalizada nos *softwares* de detecção remota, o que torna a sua aplicação difícil, sobretudo em grande escalas como em programas operacionais nacionais.

Tendo em conta as vantagens e desvantagens apresentadas pelos classificadores da máxima verosimilhança e das árvores de decisão, estes foram seleccionados para desenvolver uma metodologia de classificação automática de imagens de satélite no âmbito do projecto COSMIC. Esta opção permite conciliar a robustez e a simplicidade de um classificador largamente usado com as vantagens trazidas pelos classificadores não-paramétricos. Para além de permitir alargar as opções metodológicas, a opção de usar dois classificadores permite ainda corroborar e comprovar conclusões acerca da capacidade das imagens AWiFS em distinguir a paisagem portuguesa.

### **3.4.2. Amostragem de treino**

A recolha das observações da amostra de treino consistiu na selecção determinística de *pixels* isolados ao longo da área de estudo para evitar efeitos de autocorrelação (Campbell, 1981; Labovitz, 1986). Como estes autores demonstraram, a recolha de observações isoladas, em alternativa a outras estratégias de amostragem (e.g. recolha de observações em blocos de *pixels*) permitem obter melhores resultados na classificação assistida de imagens de satélite. Isto acontece porque a recolha de *pixels* isolados espalhados pela área a classificar abrange de forma mais eficaz a variabilidade espectral que cada classe

apresenta em função do espaço, ou seja, a amostra é mais representativa da área a classificar.

A recolha das observações teve por base a interpretação visual de ortoimagens aéreas de muito grande resolução espacial (50 cm) que cobrem a totalidade do território continental português, obtidas entre 2004 e 2006, um mosaico de imagens LISS-III e SPOT 4 de 2006, e diversa informação auxiliar (Tabela 4).

Foi seguido um esquema convencional de amostragem, isto é, foram seleccionados *pixels* puros de cada classe de ocupação do solo. Ao longo do desenvolvimento da metodologia, foi recolhido um total de 1405 observações na área de estudo (Figura 7).

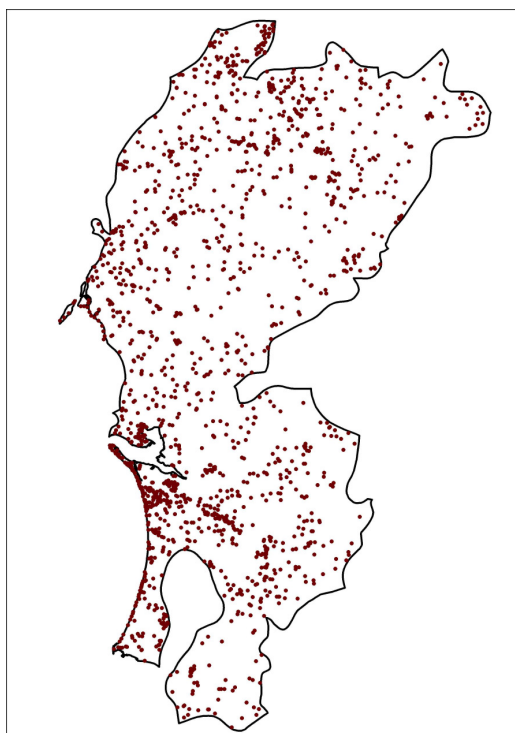


Figura 7 – Amostra de treino

A cada uma das observações foram atribuídos dois códigos. Para além do código relacionado com a respectiva classe de ocupação do solo da nomenclatura da Tabela 8, foi definido um segundo código para as observações pertencentes à mesma classe, mas que apresentassem diferentes assinaturas espectrais. Por outras palavras, o primeiro código identifica a classe de informação (classe da nomenclatura de ocupação do solo) e o segundo código identifica a classe espectral. Por exemplo, duas observações podem pertencer à mesma classe de informação (culturas de regadio), mas ter diferentes assinaturas espectrais (pertencer a diferentes classes espectrais) conforme a área a que as observações dizem respeito são exploradas com uma ou mais culturas por ano, o que determina diferentes respostas espectrais. Alguns das classes espectrais foram criadas na sequência da interacção entre as várias fases da metodologia, o que levou a vários ajustes

da amostra de treino em função de resultados anteriores. Mais concretamente, algumas classes espectrais foram criadas propositadamente para treinar áreas onde resultados anteriores demonstraram dificuldades na classificação. E além disso, consistiu num esforço para tornar a amostra de treino normalmente distribuída como o classificador da máxima verosimilhança requer (Jesen, 1996). No total, foram identificadas 37 classes espectrais (Tabela 9). Assim, o segundo código permitiu a possibilidade de usar classes espectrais em vez de classes de informação (classes de ocupação do solo) na fase de treino.

Classe espectral	Código	Classe espectral	Código
Área artificial contínua	1	Floresta de resinosas (Pinheiro bravo em dunas)	85
Área artificial descontínua	2	Áreas agro-florestais	91
Culturas de sequeiro colhidas em Julho	31	Áreas agro-florestais com solo nu	92
Culturas de sequeiro antes de Julho	32	Matos fechados	101
Pousio	33	Matos abertos	102
Culturas de regadio em início de crescimento	41	Vegetação herbácea natural mais húmida	111
Culturas de regadio no máximo de crescimento	42	Vegetação herbácea natural mais seca	112
Culturas de regadio com duas ou mais culturas por ano	43	Antigos povoamentos florestais em solo arenoso	113
Arrozais em início de crescimento	51	Recentes povoamentos florestais	114
Arrozais no máximo de crescimento	52	Cortes florestais	12
Arrozais alagados em Abril	53	Zonas húmidas	131
Vinhas	6	Sapais	132
Floresta de folhosas (Quercíneas)	71	Salinas	133
Floresta de folhosas (Eucaliptos)	72	Pedreiras e solo nu	141
Floresta de folhosas (Eucaliptos menos fechados)	73	Praias	142
Floresta de resinosas (Pinheiro manso)	81	Dunas	143
Floresta de resinosas (Pinheiro bravo)	82	Água pouco profunda	151
Floresta de resinosas (Pinheiro manso menos fechado)	83	Água	152
Floresta de resinosas (Pinheiro bravo menos fechado)	84		

Tabela 9 – Classes espectrais passíveis de serem usadas na fase de treino



### 3.4.3. Testes de classificação

No âmbito do problema da desactualização da informação sobre a ocupação do solo e dos objectivos definidos, foram realizados vários testes na área de estudo para avaliar três pontos principais:

- i) A melhor abordagem de classificação;
- ii) O efeito das características da amostragem na classificação;
- iii) A capacidade das imagens AWiFS para originar cartografia de ocupação do solo em Portugal Continental.

O primeiro ponto inclui a avaliação dos dois classificadores, o uso de probabilidades *a priori*, a melhor estratégia de classificação (numa única ou em várias fases), a integração de objectos na classificação ao nível do *pixel* (classificação híbrida) e o melhor conjunto intra-anual de imagens a serem usadas no processo de classificação.

No segundo ponto pretendeu-se avaliar a diferença entre treinar os classificadores com as classes de informação (classes de ocupação do solo do mapa) ou com classes espectrais e ainda avaliar o efeito de usar diferentes tamanhos da amostra de treino. Mais concretamente, no primeiro caso, foram realizados testes de classificação onde foram usados um maior número de classes na fase de treino (classes espectrais, referidas no ponto 3.4.2) que foram agregadas após a classificação nas suas respectivas classes de ocupação do solo. De referir que este procedimento não alterou o tamanho da amostra, mas mudou o número de observações por classe porque as 1405 observações foram separadas num maior número de classes de treino. Assim, o uso de mais classes de treino acaba por ter um efeito semelhante ao uso de diferentes tamanhos de amostra de treino, porque também, neste último caso, naturalmente diferentes tamanhos da amostra originam diferentes números de observações por classe.

Por fim, no conjunto dos dois primeiros pontos principais, bastantes importantes no que diz respeito aos objectivos no projecto COSMIC (nomeadamente, os custos de produção), é avaliada a capacidade das imagens AWiFS em originar cartografia de ocupação do solo em Portugal Continental.

Para concluir acerca destes três pontos principais, foram realizados 16 testes de classificação (Tabela 10) onde foram testados diferentes variáveis:

- i) Classificador (classificador da máxima verosimilhança e árvores de decisão);
- ii) Classes de treino (classes espectrais ou classes de ocupação do solo);
- iii) Tamanho da amostra;
- iv) Variáveis de *input* (datas, NDVI e principais componentes);
- v) Nomenclaturas de ocupação do solo (com 15 classes ou menos);

- vi) Estratégias de classificação (discriminação das classes numa única classificação, em várias classificações por fases ou numa classificação híbrida que combina a classificação ao nível do *pixel* e objectos).

No caso dos testes realizados com o classificador da máxima verosimilhança, foi usada a aplicação comercial PCI Geomatics®. As árvores de decisão foram derivadas pelo See5®, que tem a limitação de ser um *software* que não efectua processamento de imagens. Assim, a classificação das imagens foi realizada com recurso ao Microsoft® Access, que estabeleceu a ligação entre o See5® e um Sistema de Informação Geográfica (ArcGIS 9.2®). Os testes de classificação com árvores de decisão aqui apresentados usaram 25% de *pruning* e 10 árvores de *boosting*. Na Tabela 10 estão resumidos todos os testes realizados e na Tabela 11 estão resumidas as avaliações que a comparação entre os testes de classificação permitem fazer acerca da influência das variáveis referidas acima. Finalmente, os objectivos atribuídos a cada um dos testes e outras considerações são apresentadas nos próximos parágrafos.

Teste	Classificador	Classes de treino	Classes do mapa	Variáveis de <i>input</i>	Estratégia de classificação	Tamanho da amostra de treino
1	CMV e AD	37	15	Abril e Julho	Única	Inteira
2	CMV e AD	37	15	Abril, Julho e Outubro	Única	Inteira
3	CMV e AD	15	15	Abril e Julho	Única	Inteira
4	CMV e AD	15	15	Abril, Julho e Outubro	Única	Inteira
5	CMV e AD	15	15	Abril, Julho e Outubro, NDVI de Abril e Julho	Única	Inteira
6	CMV e AD	15	15	CP de Abril e Julho	Única	Inteira
7	CMV e AD	15	15	CP das três datas	Única	Inteira
8	CMV com prob. <i>a priori</i>	15	15	CP de Abril e Julho	Única	Inteira
9	CMV e AD	10	10	Abril, Julho e Outubro	Única	Inteira
10	CMV e AD	6	6	Abril, Julho e Outubro	Única	Inteira
11	CMV e AD	15	10	Abril, Julho e Outubro	Única	Inteira
12	CMV e AD	15	6	Abril, Julho e Outubro	Única	Inteira
13	AD	15	15	Abril, Julho e Outubro	Única	5 amostras aleatórias com metade do tamanho
14	AD	15	15	Abril; Julho; Outubro	Por fases	Inteira
15	AD	15	15	Abril, Julho e Outubro	Por fases	Inteira
16	CMV/Objecto	37	15	Abril e Julho	Híbrida	Inteira

Tabela 10 – Características dos testes de classificação

Comparação entre testes	Avaliação					
	Classif.	Variáveis de <i>input</i>	Classes de treino ou tamanho da amostra	Estratégia de classificação	Nomenclatura de ocupação do solo	<i>Pixel</i> vs. híbrida
1 a 12	x					
1/2 e 3 a 7		x				
1/3, 2/4, 9/11, 10/12 e 4/13			x			
4/14 e 4/15				x		
4/9/10					x	
1 e 16						x

Tabela 11 – Avaliações permitidas através da comparação de testes de classificação

Os testes de classificação 1 a 5 foram pensados para avaliar, simultaneamente, o impacto de treinar ambos os classificadores com as classes de informação ou com as classes espectrais, assim como para avaliar o melhor conjunto intra-anual de imagens para serem usadas como variáveis de *input*. Desta forma, os testes 1 a 5 diferem acerca das classes de treino e/ou das variáveis de *input*. Os testes 1 e 2 usaram 37 classes espectrais para treinar os classificadores (posteriormente agrupadas na nomenclatura de 15 classes), diferenciando entre si nas variáveis de *input* com Abril e Julho ou as três datas, respectivamente. Os testes 3 e 4 usaram as mesmas variáveis de *input* que os testes anteriores, mas usaram 15 classes da nomenclatura na fase de treino. A decisão de testar a classificação apenas com imagens relativas a duas datas teve por objectivo verificar se a perda de informação espectral se traduziria em perdas acentuadas de exactidão da classificação, o que é relevante em termos de custos operacionais relacionados com a aquisição de imagens. A selecção das imagens de Abril e Julho teve por base resultados de testes preliminares que revelaram que esta combinação proporcionava melhores resultados que qualquer outra combinação. No teste 5 pretendeu-se verificar a mais-valia de acrescentar o índice de vegetação NDVI à informação espectral disponível. Este teste apenas foi realizado com 15 classes de ocupação do solo devido a problemas relacionados com a dimensionalidade dos dados revelados nos testes anteriores, aquando do uso das 37 classes espectrais.

Nos testes de classificação 6 e 7 pretendeu-se avaliar a exactidão da classificação obtida com a redução da dimensionalidade das variáveis de *input*. Esta opção optimiza a implementação do classificador (Lillesand e Kiefer, 1994) e com ela pretendeu-se ultrapassar alguns problemas revelados pelos resultados dos testes anteriores. Neste sentido, procedeu-se ao cálculo das componentes principais (CP) usando as imagens de Julho e Abril (teste 6) e das três datas (teste 7) através da aplicação comercial ArcGIS 9.2©. No primeiro caso, foram usadas 5 componentes que explicam 98.7% da informação contida

nas duas datas e no segundo caso foram usadas 7 componentes que explicam 99.24% da informação contida nas três datas.

O teste 8 consistiu numa alteração ao CMV. A partir do teste 6, que produziu o melhor resultado (ver ponto 4), as imagens foram classificadas usando probabilidades *a priori*. As probabilidades foram calculadas através da percentagem de área ocupada por cada classe na área de estudo segundo o CORINE Land Cover 2000 (Tabela 12). O CMV usado nos testes anteriores calcula a probabilidade dos indivíduos pertencerem a cada uma das classes, cuja probabilidade de ocorrerem é igual para todas. Daqui pode resultar que as classes mais frequentes sejam sobclassificadas e as classes menos frequentes sejam sobreclassificadas. Assim, a inclusão de probabilidades *a priori* teve por objectivo reduzir erros de classificação através do favorecimento das classes mais frequentes e do desfavorecimento das classes menos frequentes com base na frequência esperada das mesmas.

Classe de ocupação do solo	Probabilidade	Classe de ocupação do solo	Probabilidade
1	0.008	9	0.109
2	0.010	10	0.029
3	0.109	11	0.005
4	0.072	12	0.008
5	0.037	13	0.004
6	0.028	14	0.002
7	0.439	15	0.015
8	0.125		

Tabela 12 – Probabilidades *a priori* usadas no teste 8 e calculadas a partir do CLC2000

Os testes de classificação 9 a 12 foram definidos para avaliar duas questões. Por um lado, para avaliar o potencial melhoramento da exactidão da classificação ao usar uma nomenclatura de ocupação do solo menos detalhada. Assim, nos testes 9 e 11 foi usada uma nomenclatura constituída por 10 classes (Tabela 13) que derivou da nomenclatura apresentada na Tabela 8. No mesmo sentido, nos testes 10 e 12 foi usada uma nomenclatura constituída por 6 classes (Tabela 13).

Código	Nomenclatura com 15 classes (da Tabela 8)	Código	Nomenclatura com 10 classes	Código	Nomenclatura com 6 classes
1	Áreas artificiais contínuas	1	Áreas artificiais	1	Áreas artificiais
2	Áreas artificiais descontínuas	2	Culturas de sequeiro	2	Agricultura
3	Culturas de sequeiro	3	Culturas irrigadas	3	Floresta
4	Culturas de regadio	4	Vinhas	4	Áreas naturais não florestais
5	Arrozais	5	Floresta	5	Solo nu
6	Vinhas	6	Áreas agro-florestais	6	Água e zonas húmidas
7	Florestas de folhosas	7	Áreas naturais não florestais		
8	Florestas de resinosas	8	Cortes florestais		
9	Áreas agro-florestais	9	Solo nu		
10	Matos	10	Água e zonas húmidas		
11	Vegetação herbácea natural				
12	Cortes florestais				
13	Zonas húmidas				
14	Solo nu				
15	Água				

Tabela 13 – Nomenclaturas de ocupação do solo com 15, 10 e 6 classes

Por outro lado, os testes 9 a 12 foram também definidos para avaliar o efeito de treinar ambos os classificadores com classes de informação ou com classes espectrais (que são depois agrupadas pós-classificação). Assim, estes testes são semelhantes aos testes 1 a 4 (onde os classificadores foram treinados com 37 ou 15 classes de treino). Mas neste caso, nos testes 11 e 12, a nomenclatura com 15 classes de ocupação do solo não foi usada como classes de informação, mas sim usada como classes espectrais, e posteriormente à classificação foram agrupadas na nomenclatura da cartografia (10 e 6 classes de ocupação do solo, respectivamente). Os testes 11 e 12, juntamente com os testes 1 a 4, são importantes para avaliar o esforço de amostragem desenvolvido e o tempo despendido na recolha e identificação das observações que é necessário para a classificação produzir resultados satisfatórios.

O teste 13 foi definido para explorar mais detalhadamente o efeito do tamanho da amostra de treino no desempenho das árvores de decisão, uma vez que se encontram relatos

contraditórios acerca desta matéria na literatura especializada. Neste teste, foram realizadas cinco classificações com metade do tamanho da amostra (metade das observações por classe), seleccionadas aleatoriamente da amostra de treino para avaliar a exactidão da classificação com um menor esforço de amostragem.

Nos testes 14 e 15 foram exploradas duas estratégias de classificação diferentes dos testes anteriores. Nestes dois testes, a estratégia consistiu na produção de cartografia através de várias classificações por fases. O teste 14 incluiu três classificações independentes, cada uma com base na informação espectral de cada uma das datas das imagens AWiFS (Abril, Julho e Outubro) e com nomenclaturas de ocupação do solo diferentes (Tabela 14). Para cada data, a respectiva nomenclatura usada neste teste foi definida através da união das classes que entre si apresentam a mesma ocupação do solo naquela data. Como exemplo, pode-se referir que a classe Vinha e Solo nu podem apresentar a mesma resposta espectral nos meses de primavera devido ao facto de as vinhas, durante esta altura, não apresentarem qualquer folhagem e o sobcoberto ser, muitas vezes, limpo, o que deixa o solo sem vegetação. Este procedimento permite reduzir o número de classes de ocupação do solo em cada data com o objectivo de reduzir os erros de classificação. As classes que são distintas entre si nas três datas, não foram alteradas (e.g. Áreas urbanas contínuas).

Classes de ocupação do solo		
Abril	Julho	Outubro
Áreas artificiais contínuas	Áreas artificiais contínuas	Áreas artificiais contínuas
Áreas artificiais descontínuas	Áreas artificiais descontínuas	Áreas artificiais descontínuas
Vegetação herbácea verde	Culturas de regadio	Vegetação herbácea verde
Vegetação herbácea pouco verde	Arrozais	Vegetação herbácea seca
Vegetação herbácea seca	Vegetação herbácea seca	Vinhas
Floresta de folhosas	Vinhas	Floresta de folhosas
Floresta de resinosas	Floresta de folhosas	Floresta de resinosas
Áreas agro-florestais	Floresta de resinosas	Áreas agro-florestais
Matos	Áreas agro-florestais	Matos
Zonas húmidas	Matos	Zonas húmidas
Solo nu	Zonas húmidas	Solo nu
Água	Solo nu	Água
Água		

Tabela 14 – Nomenclaturas de ocupação do solo menos detalhadas usadas no teste 14

Depois de realizadas as três classificações, cada *pixel* foi classificado com base na combinação das respectivas três classificações de cada data. Por exemplo, um *pixel* classificado no mês de Abril como Vegetação herbácea verde e nos meses de Julho e Outubro como Vegetação herbácea seca foi classificado como Cultura de sequeiro. A

Tabela 15 mostra as combinações das classificações que dão origem à nomenclatura final de 15 classes, realçando-se (em cor azul) as classes que resultam da combinação de diferentes ocupações do solo nas três datas.

Classes de ocupação do solo			Classes de ocupação do solo do mapa
Abril	Julho	Outubro	
Áreas artificiais contínuas	Áreas artificiais contínuas	Áreas artificiais contínuas	Áreas artificiais contínuas
Áreas artificiais descontínuas	Áreas artificiais descontínuas	Áreas artificiais descontínuas	Áreas artificiais descontínuas
Vegetação herbácea verde	Vegetação herbácea seca ou Solo nu	Vegetação herbácea seca	Culturas de sequeiro
Solo nu, Vegetação herbácea verde ou seca	Vegetação herbácea verde	Vegetação herbácea seca ou verde	Culturas de regadio
Solo nu, Vegetação herbácea seca ou Água	Arrozais	Vegetação herbácea seca	Arrozais
Solo nu ou Vegetação herbácea seca	Vinhas	Vegetação herbácea seca	Vinhas
Florestas de folhosas	Florestas de folhosas	Florestas de folhosas	Florestas de folhosas
Florestas de resinosas	Florestas de resinosas	Florestas de resinosas	Florestas de resinosas
Áreas agro-florestais	Áreas agro-florestais	Áreas agro-florestais	Áreas agro-florestais
Matos	Matos	Matos	Matos
Vegetação herbácea pouco verde	Vegetação herbácea seca	Vegetação herbácea seca	Vegetação herbácea natural
Florestas de folhosas ou de resinosas	Solo nu, Florestas de folhosas ou de resinosas	Solo nu	Cortes florestais
Zonas húmidas	Zonas húmidas	Zonas húmidas	Zonas húmidas
Solo nu	Solo nu	Solo nu	Solo nu
Água	Água	Água	Água

Tabela 15 – Esquema de classificação por fases do teste 14

No teste 15 foi explorada outra estratégia de classificação por fases. As observações de todas as classes de ocupação do solo estavam juntas no início do teste e foram posteriormente separadas por diferentes árvores de decisão. A primeira fase consistiu na construção de uma árvore de decisão que separou as classes Água e Zonas húmidas das restantes classes. Na segunda fase, por um lado, as classes Água e Zonas húmidas foram separadas, e por outro lado, as restantes classes foram separadas em classes com e sem

vegetação. Este processo continuou sucessivamente até as 15 classes serem distinguidas (Figura 8).

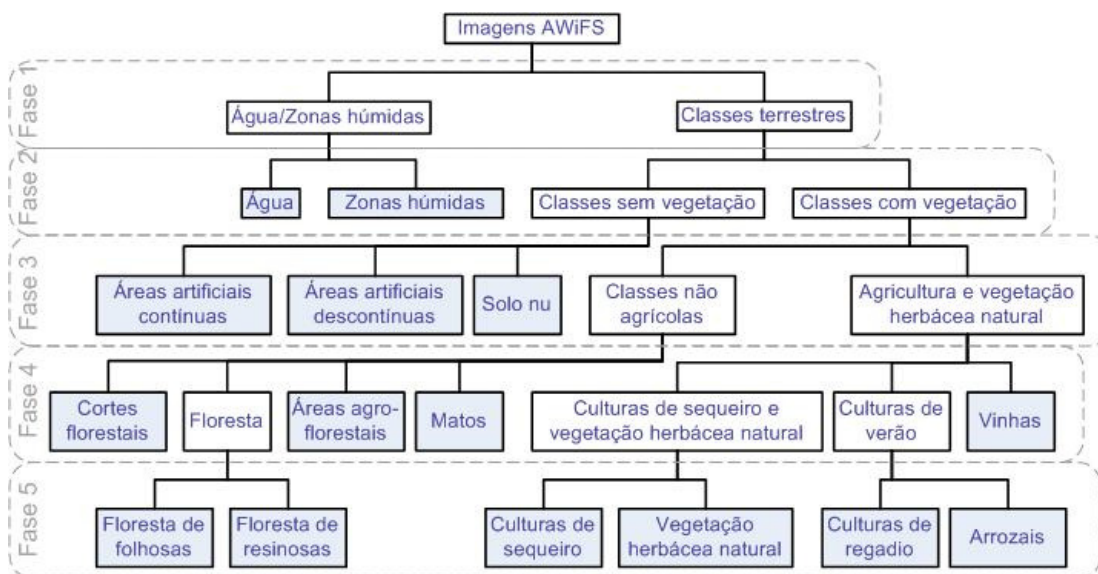


Figura 8 – Classificação por fases desenvolvida no teste de classificação 15

Por fim, o ultimo teste de classificação (teste 16) teve como objectivo fazer uma avaliação exploratória da potencialidade de usar a análise orientada por objectos, o que constitui uma abordagem mais complexa e ainda não usada em programas operacionais de grande escala. Assim, com recurso ao trabalho desenvolvido ao nível do *pixel*, realizou-se uma classificação híbrida com a integração de objectos extraídos das imagens AWiFS. Os objectos foram delimitados com recurso ao *software* eCognition© através da segmentação das imagens AWiFS de Abril, Julho e Outubro. Foram realizadas várias experiências de onde se chegou à conclusão que eram adequados os seguintes parâmetros de segmentação: *color*, 0.8, *shape*, 0.5, *smoothness*, 0.5 e *compactness*, 0.5. Com estes parâmetros, a segmentação das imagens deram origem a 88356 objectos, em média com 14.51 ha e desvio padrão de 18.18 ha. Posteriormente, o resultado da segmentação foi sobreposto à classificação ao nível do *pixel* e cada objecto foi classificado na classe que ocupava a maior percentagem de área. Uma vez que cada um dos resultados dos testes 1 a 15 poderia ser usado na classificação híbrida para classificar os objectos, foi seleccionado o resultado do teste 1 com o CMV a título de exemplo e por ser um teste que apresentou um dos melhores resultados.

### 3.5. Avaliação da exactidão do mapa

Todos os resultados dos testes de classificação foram sujeitos a uma avaliação da sua qualidade através de um processo rigoroso e estatisticamente válido, o que permite derivar índices de qualidade temática não enviesados e produzir uma avaliação da qualidade defensável do ponto de vista científico (Stehman e Czaplewski, 1998). O processo de



validação consistiu na construção de uma amostra da ocupação do solo da área de estudo com a qual cada resultado foi comparado.

Como informação de referência usaram-se ortoimagens aéreas de muito grande resolução espacial (50 cm) que cobrem a totalidade do território continental português, obtidas entre 2004 e 2006, e um mosaico de imagens LIS-III e SPOT 4 de 2006 (Tabela 4).

Para a recolha das observações da amostra de validação, foi definida uma área de 60 por 60 m (3600m<sup>2</sup>), equivalente a um *pixel* das imagens AWiFS, como unidade de amostragem. A selecção do conjunto das unidades de amostragem (observações) realizou-se através de uma amostragem aleatória estratificada por classe de ocupação do solo a partir de um mapa preliminar obtido durante os testes. Esta opção deveu-se ao facto desta estratégia de amostragem garantir que todas as classes são amostradas. Então, recolheram-se aleatoriamente as observações por cada classe de ocupação do solo do mapa preliminar. Seria indicado recolher 100 unidades de amostragem por estrato visto que este número de observações por classe garante que a estimativa da exactidão global do mapa e específica das classes tenha uma precisão absoluta nunca superior a 10%, para um nível de significância de 0.05 (Stehman, 2001). No entanto, a recolha de 100 observações por classe seria um trabalho moroso, não compatível com o processo de desenvolvimento dos testes de classificação. Assim, definiu-se um compromisso entre a precisão da estimativa dos índices de qualidade e a rapidez do processo, ao recolher 30 observações por classe, o que garante uma precisão absoluta não superior a 18.26% (Figura 9).

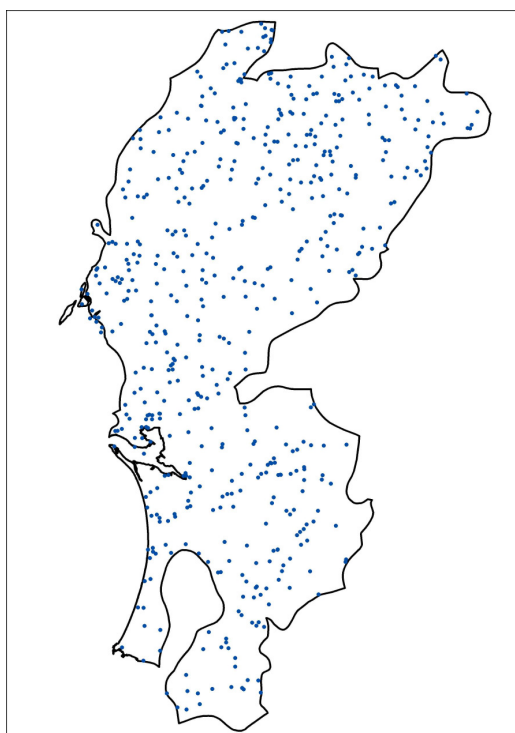


Figura 9 – Amostra de validação

Seria igualmente mais indicado que, para cada resultado dos testes de classificação, se procedesse à recolha de observações para a amostra de validação de acordo com a estratificação produzida por esse resultado. Assim, poder-se-ia controlar e igualar os valores de precisão absoluta da estimativa dos índices de qualidade de todos os resultados. Mas mais uma vez, este seria um trabalho muito moroso. Assim, a adopção da mesma estratificação proveniente do mapa preliminar na validação de todos os resultados foi uma solução de compromisso que se considerou adequada. Desta forma, foi possível canalizar mais esforços para o desenvolvimento e teste de uma aproximação metodológica a aplicar no projecto COSMIC, considerada, nesta fase, prioritária. Uma vez que os mapas a validar são diferentes daquele que foi utilizado para a estratificação, é de esperar que as classes apresentem um número diferente de observações por classe. Isto significa que cada uma destas classes irá ver a sua estimativa da exactidão calculada com diferentes máximos para as precisões absolutas. Contudo, espera-se que estas estimativas usem máximos de precisão absoluta não muito distintos, admitindo que o mapa que deu origem aos estratos é um mapa aproximado aos resultados dos testes de classificação.

Após a selecção das observações da amostra de validação, a informação de referência foi interpretada visualmente e a cada observação foi atribuído um código principal de ocupação do solo da nomenclatura e ainda um código alternativo, quando necessário. Esta necessidade relaciona-se com dois aspectos. Por um lado, com os *pixels* de transição ou mistos entre duas ou mais classes de ocupação do solo (e.g. uma observação que se localize na linha de fronteira entre uma parcela de culturas de sequeiro e de regadio); por outro lado, com a incerteza na atribuição da classe de ocupação do solo a algumas observações que contêm características de mais do que uma classe (e.g. uma área urbana cuja densidade de artificialização é intermédia entre os casos típicos de áreas artificiais contínuas e descontínuas, pelo que se considera aceitável ser classificada numa ou noutra classe).

Nesta fase, a diferente abordagem de classificação experimentada no teste 16 (classificação híbrida) obrigou a um procedimento diferente. Neste caso, a unidade de amostragem não corresponde ao *pixel*, mas sim ao objecto. Mais uma vez, seria muito moroso proceder a uma avaliação independente da exactidão do mapa resultante. Assim, as observações da amostra de validação do teste 16 (objectos) foram seleccionadas através da sobreposição das observações da amostra de validação usada nos testes anteriores (*pixels*). Este procedimento produziu uma amostra de validação inferior porque alguns objectos do teste 16 incorporaram mais do que uma observação dos testes anteriores. Assim, às 570 observações da amostra de validação usada nos testes 1 a 15 corresponderam a 565 observações usadas na amostra de validação do teste 16. Posteriormente, a cada observação foi atribuído um novo código principal e alternativo de ocupação do solo, agora relativamente à ocupação do solo maioritária em todo o objecto.

Depois de construídas as amostras de validação, foi construída a matriz de confusão para cada teste de classificação e foram avaliados através dos índices de qualidade global e específica das classes da cartografia produzida. Cada observação da amostra de validação foi considerada bem classificada se um dos dois códigos (principal ou alternativo) correspondia com a classe produzida pelos classificadores. Caso nenhum dos códigos de referência concordasse com a classe do mapa, é o código principal que consta na matriz.

Nesta fase, a matriz de confusão foi construída através da contagem das observações bem e mal classificadas (Figura 1), devido à simplicidade da sua execução e sua ampla aplicação. No entanto, esta forma de construir a matriz de confusão não considera a diferença de área ocupada por cada classe de ocupação do solo. Como a construção da amostra de validação seguiu uma amostragem estratificada por classe de ocupação do solo, existem diferenças na intensidade de amostragem de cada classe, de onde resulta a introdução de enviesamento (Card, 1982).

Posteriormente, para os melhores testes de classificação de cada classificador, foi construída outra matriz de confusão ponderada com área ocupada por cada classe de ocupação do solo. A construção desta matriz calcula os índices de qualidade do mapa, não com base na proporção de amostras bem classificadas, mas sim com base na proporção de área bem classificada. As equações utilizadas para derivar os índices de qualidade específica e global foram as propostas por Carrão *et al.* (2007), derivadas do trabalho de Cochran (1977), e que se passam a apresentar de seguida. Dada uma matriz de confusão e supondo que:

$N$  – número de *pixels* no mapa;

$N_h$  – número de *pixels* na categoria  $h$  do mapa;

$N_{h1}$  – número de *pixels* na categoria  $h$  do mapa que correspondem à mesma categoria na referência;

$N_g$  – número de *pixels* na categoria  $g$  da referência;

$n_h$  – número de *pixels* da amostra recolhidos na categoria  $h$  do mapa;

$n_{hg}$  – número de *pixels* da amostra recolhidos na categoria  $h$  do mapa que intersectam a categoria  $g$  da referência;

$n_{h1}$  - número de *pixels* da amostra recolhidos na categoria  $h$  do mapa que correspondem à mesma categoria na referência.

A equação para a estimação da exactidão global (EG) do mapa, isto é, a proporção de *pixels* correctamente classificados no mapa é:

$$\hat{P}_c = \frac{1}{N} \sum_{h=1}^H \frac{N_h}{n_h} n_{h1}$$

e a respectiva variância estimada é:

$$\hat{V}(\hat{P}_c) = \sum_{h=1}^H \left( \frac{N_h}{N} \right)^2 \frac{N_h - n_h}{N_h n_h} \left[ \frac{n_{h1}}{n_h} \left( 1 - \frac{n_{h1}}{n_h} \right) \right]$$

A equação para a estimação da exactidão do utilizador (EP) para cada classe do no mapa é:

$$\hat{P}_{h,c} = \frac{n_{h1}}{n_h}$$

e a respectiva variância estimada é:

$$\hat{V}(\hat{P}_{h,c}) = \left( 1 - \frac{n_h}{N_h} \right) \frac{1}{n_h - 1} \left[ \frac{n_{h1}}{n_h} \left( 1 - \frac{n_{h1}}{n_h} \right) \right]$$

Para a estimação da exactidão do produtor (EP), num processo decorrente de uma amostragem aleatória estratificada, deve ser calculado utilizando uma estimação não enviesada das quantidades desconhecida da população  $N_{h1}$  e  $N_g$ . Os seus estimadores não enviesados são:

$$\hat{N}_{h1} = \frac{N_h}{n_h} n_{h1}$$

$$\hat{N}_g = \sum_{h=1}^H \frac{N_h}{n_h} n_{hg}$$

A exactidão do produtor pode ser vista como o quociente destes dois estimadores de acordo com a seguinte equação:

$$\hat{P}_{g,c} = \frac{\hat{N}_{h1}}{\hat{N}_g}$$

e o respectivo Erro Médio Quadrático (EMQ) estimado é:

$$EMQ(\hat{P}_{g,c}) = \left( \frac{N}{\hat{N}_g} \right)^2 \sum_{h \neq g} \left[ \left( \frac{N_h}{N} \right)^2 \left( \frac{N_h - n_h}{N_h n_h} \right) \frac{n_{hg}}{n_h} \left( 1 - \frac{n_{hg}}{n_h} \right) \hat{P}_{g,c}^2 \right] + \left[ \left( \frac{N_h}{N} \right)^2 \left( \frac{N_h - n_h}{N_h n_h} \right) \frac{n_{hg}}{n_h} \left( 1 - \frac{n_{hg}}{n_h} \right) (1 - \hat{P}_{g,c}^2)^2 \right]$$

A partir das variâncias é possível calcular a precisão absoluta  $d$  para cada um dos  $\hat{P}$  estimados (EU, EP e EG), sendo que  $d = Z_{\alpha/2} [V(\hat{P})]^{1/2}$ , onde  $Z_{\alpha/2}$  é o quantil da distribuição normal para um nível de confiança de  $1-\alpha$ , com  $\alpha = 0,05$ , no presente caso.

## 4. Resultados e discussão

Os vários testes de classificação produziram um conjunto de resultados que devem ser discutidos individualmente e em conjunto. Desta forma, nos próximos capítulos são apresentados e discutidos os resultados globais de cada teste de classificação (4.1), a influência das variáveis dos testes de classificação (4.2), as matrizes de confusão ponderada com a área classificada dos melhores testes de cada classificador (ao nível do *pixel*) (4.3), a sua comparação com o mapa CLC2000 (4.4) e ainda as imagens AWiFS (4.5). Por fim, são dirigidas críticas e perspectivas futuras ao trabalho desenvolvido (4.6).

### 4.1. Exactidão global dos testes de classificação

Os valores de exactidão global dos testes de classificação são apresentados na Tabela 16.

Teste	Exactidão Global (%)		Classes de treino	Classes do mapa	Variáveis de <i>input</i>	Estratégia de classificação	Tamanho da amostra de treino
	CMV	AD					
1	62.11	55.96	37	15	Abril e Julho	Única	Inteira
2	37.72	55.96	37	15	Abril, Julho e Outubro	Única	Inteira
3	61.58	57.54	15	15	Abril e Julho	Única	Inteira
4	59.30	56.67	15	15	Abril, Julho e Outubro	Única	Inteira
5	53.51	58.98	15	15	Abril, Julho e Outubro, NDVI de Abril e Julho	Única	Inteira
6	62.98	60.53	15	15	CP de Abril e Julho	Única	Inteira
7	60.35	58.68	15	15	CP das três datas	Única	Inteira
8 <sup>1</sup>	58.25	-	15	15	CP de Abril e Julho	Única	Inteira
9	71.58	72.1	10	10	Abril, Julho e Outubro	Única	Inteira
10	74.74	78.4	6	6	Abril, Julho e Outubro	Única	Inteira
11	67.37	63.86	15	10	Abril, Julho e Outubro	Única	Inteira
12	71.40	67.54	15	6	Abril, Julho e Outubro	Única	Inteira
13	-	58.25	15	15	Abril, Julho e Outubro	Única	5 amostras aleatórias com metade do tamanho
	-	58.42					
	-	56.84					
	-	55.26					
	-	59.82					
14	-	54.03	15	15	Abril; Julho; Outubro	Por fases	Inteira
15	-	62.11	15	15	Abril, Julho e Outubro	Por fases	Inteira
16 <sup>2</sup>	67.26	-	37	15	Abril e Julho	Híbrida	Inteira

1 – Teste que integrou probabilidades *a priori*

2 – Classificação híbrida com a combinação de objectos e o mapa do teste 1 com o CMV

Tabela 16 – Resultados e características dos testes de classificação

Como se pode verificar na Tabela 16, os testes 1 a 5 demonstram que o classificador da máxima verosimilhança é bastante sensível ao tamanho da amostragem de treino, ou por outras palavras, é bastante sensível ao número de observações de treino por classe. É oportuno lembrar que todos estes testes usaram uma amostra de treino com o mesmo tamanho (1405 observações), mas usaram diferentes classes de treino, o que provocou que houvesse diferentes números de observações por classe. Como referido no ponto 3.4.3, este procedimento é similar a usar diferentes tamanhos de amostragem. Ao usar 37 classes de treino fez com que a exactidão global do teste 1 fosse 62.11% enquanto no teste 2 (classificado com mais os dados espectrais de Outubro) fosse 37.72%. Ou seja, com maior informação espectral, ao contrário do que seria de esperar, o teste 2 devolveu um resultado muito pior que o teste 1. Este facto pode ser explicado pelo *fenómeno de Hughes*, também conhecido pela maldição da dimensionalidade dos dados. Este fenómeno é também a razão por que o teste 2 e 4 devolveram resultados tão diferentes. No teste 4, onde se usou o CMV com a mesma informação espectral que o teste 2, mas com menos classes de treino (logo, maior número de observações por classe), o resultado obtido foi muito superior (59.30%). Isto demonstra que o tamanho da amostra de treino é insuficiente para simultaneamente treinar o CMV com as 37 classes espectrais e usar mais do que duas datas (8 bandas) nas as variáveis de *input*.

Tendo em atenção os resultados anteriores, em função da dimensionalidade dos dados e do número de observações por classe de treino, seria então de esperar que o teste 3 (treinado com 15 classes) produzisse um mapa mais exacto do que o teste 1 (treinado com 37 classes). Contudo, isto coloca outra questão. O comportamento do CMV assume a distribuição dos dados como normal, mas ao treinar o CMV com as 15 classes (que não têm uma distribuição perfeitamente normal) provoca-se uma contrariedade. Assim, a exactidão conseguida com o teste 3 não superou a exactidão global do teste 1. De qualquer maneira, mesmo com o maior número de observações por classe usadas aquando do treino com 15 classes, o tamanho da amostra continua insuficiente. Isto é demonstrado pelos resultados obtidos nos testes 3 a 5, onde a exactidão global diminuiu à medida que as variáveis de *input* foram aumentando (61.58%, 59.30% e 53.51%, respectivamente). Estes primeiros testes revelam que o CMV é muito dependente da relação entre o número de observações por classe de treino e o número de as variáveis de *input*.

Por outro lado, nos testes 1 a 5, os resultados dos testes de classificação obtidos com as árvores de decisão foram similares independentemente do uso de 15 ou 37 classes de treino. Isto mostra que a construção das árvores não é afectada pelo número de observações por classe. Além disto, a produção das árvores de decisão com 15 classes de treino produziu ligeiras subidas da exactidão das classificações. Isto sugere que as árvores de decisão não requerem dados caracterizados pela distribuição normal. Mais ainda, a comparação do resultado do teste 5 com os testes 3 e 4 revela que a adição de mais

variáveis de *input* (nomeadamente, a data de Outubro e os NDVI's) proporcionaram melhorias nos resultados, embora ligeiras. A semelhança de resultados obtidos através das árvores de decisão com classes de informação ou classes espectrais vão ao encontro dos autores que referem este classificador como robusto e pouco exigente relativamente as características da amostra de teste (e.g. Friedl e Brodley, 1997; Pal e Mather, 2003).

Na consequência dos resultados dos primeiros cinco testes, os testes 6 e 7, definidos para reduzir a dimensionalidade dos dados, permitiram evitar o *fenómeno de Hughes* observado entre os resultados dos testes 1 e 2 com o CMV. Os testes 6 e 7, que diferem dos testes 1 e 2 em relação às variáveis de *input* e classes de treino, permitiram evitar o efeito da maldição da dimensionalidade dos dados e revelam ligeiras melhorias nos resultados da classificação efectuadas com ambos os classificadores. As componentes principais diminuíram a dimensionalidade dos dados e removeram a informação redundante das bandas originais das imagens AWiFS. Apesar disso, a exactidão global destes testes não subiu muito, o que sugere que as imagens não contêm informação suficiente para a obtenção de melhores resultados.

O uso de probabilidades *a priori* no teste 8 não surtiu melhorias na exactidão da classificação. Parte da diminuição do valor de exactidão global poderá derivar de as probabilidades terem sido calculadas a partir do CLC2000, de duas formas. Este é um produto cartográfico com uma unidade mínima cartográfica muito superior à resolução das imagens AWiFS e a nomenclatura não é directamente comparável com todas as classes da nomenclatura usada, o que poderá afastar a frequência de algumas classes da sua real proporção.

Os testes 9 a 12 foram pensados para avaliar a adequabilidade das imagens AWiFS para produzir cartografia de ocupação do solo menos detalhada tematicamente. Naturalmente, os resultados dos testes de classificação, com ambos os classificadores, foram mais elevados com uma nomenclatura do solo menos detalhada e obtiveram-se resultados superiores a 70% (Tabela 16). Contudo, a impossibilidade de obter resultados ainda mais elevados significa que existe confusão entre classes de ocupação do solo pouco semelhantes. Além disso, assim como os testes anteriores, estes demonstram que a definição da fase de treino é bastante importante para a exactidão da classificação. Ao treinar os classificadores com mais classes do que o número de classes final da nomenclatura do mapa (testes 11 e 12) não produziram melhores resultados, antes pelo contrário, pioraram (comparando com os testes 9 e 10). No caso do CMV, pode ser argumentado que a dimensionalidade dos dados é o responsável por este decréscimo da exactidão (à semelhança do decréscimo observado entre os testes 4 e 2), mas o mesmo não é razoável para o caso das árvores de decisão. Ao comparar os testes 1 a 4 com os testes 9 a 12, é possível observar que as árvores de decisão alcançaram resultados mais elevados com um treino mais heterogéneo, i.e., quando as classes da nomenclatura do mapa não estavam separadas em classes

espectrais na fase de treino. Assim, o uso de classes espectrais na fase de treino, posteriormente agrupadas nas classes de ocupação do solo, parece não ser um procedimento que aumente a exactidão da classificação. Este facto é bastante importante porque a definição de classes espectrais é um processo demorado.

O teste 13 foi definido para explorar de uma forma mais aprofundada o efeito do tamanho da amostra de treino na construção das árvores de decisão. Á imagem dos testes 1 a 5, este teste reforça que as árvores de decisão não são sensíveis ao tamanho da amostra e a exactidão da classificação não é afectada. De uma forma geral, as cinco classificações realizadas com metade do tamanho da amostra de treino (metade das observações por classe) não resultaram em valores de exactidão global mais baixos do que o teste 4 (treinado com as mesmas condições mas com a totalidade da amostra de treino) e produziram valores de exactidão global inferiores a 60%. A semelhança de resultados obtidos através das árvores de decisão com classes de informação ou classes espectrais e com metade do tamanho da amostra de treino vão ao encontro dos autores que referem este classificador como robusto e pouco exigente relativamente as características da amostra de treino (e.g. Friedl e Brodley, 1997; Pal e Mather, 2003).

Os testes 14 e 15, os quais foram definidos para avaliar diferentes estratégias de classificação (classificação por fases), conduziram a diferentes resultados. O teste 14 produziu um resultado de 54.03% de exactidão global, isto é, pior do que o teste realizado numa única classificação com as mesmas variáveis de *input* e nomenclatura (teste 4). Este resultado deve-se à complexidade da estratégia. Após cada classificação, para cada *pixel* é decidida a classificação final considerando o perfil espectral de cada classe de ocupação do solo ao longo das três datas. Esta estratégia torna-se complexa quando erros de classificação ocorrem em pelo menos uma data. Estes erros alteram o perfil espectral dos *pixels* (Tabela 15) e provocam erros no mapa final. Assim, a estratégia da classificação por fases desenvolvida no teste 14 não traz melhorias à classificação.

Por outro lado, no teste 15 foi obtido 62.11% de exactidão global, um resultado mais elevado relativamente ao teste 4, embora diminuto. Este resultado demonstra que, nesta estratégia, cada árvore de decisão conseguiu distinguir melhor as classes de ocupação do solo no conjunto das fases de classificação. E para além do resultado obtido com 15 classes, permite produzir mais mapas, um por cada uma das cinco fases de classificação (Figura 8). Desta forma, foi possível produzir mapas com 2, 4, 7, 12 e 15 classes e observar a evolução da exactidão da classificação ao longo da estratégia (Figura 10).



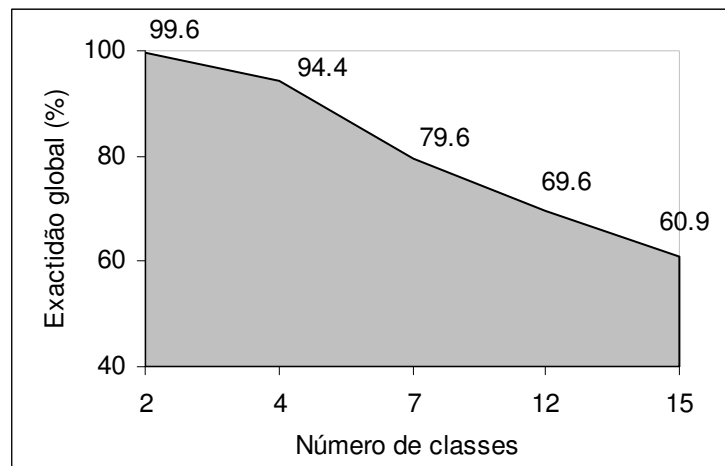


Figura 10 – Evolução da exactidão temática da estratégia de classificação por fases do teste 15

A análise da Figura 10 demonstra que até quatro classes (Água, Zonas húmidas, classes sem vegetação e classes com vegetação), a classificação é bastante exacta (94.4%) e os erros de classificação começam a ser maiores aquando da separação das classes sem e com vegetação. Através da Figura 8 é possível ver que classes de ocupação do solo correspondem a cada fase de classificação, ou então na Tabela 5.1 do Anexo 5, onde podem ser consultadas mais facilmente. De realçar a melhoria acentuada que este teste revela com a perda de pouco detalhe temático. Com 12 classes, o resultado sobe para 69.6%. O inconveniente que se poderá colocar em relação à nomenclatura da 4ª fase de classificação é a agregação de classes semelhantes em termos de ocupação do solo, mas com diferentes importâncias na perspectiva do uso da cartografia por parte do utilizador. Por este motivo, a produção de cartografia requer, normalmente, nomenclaturas o mais detalhadas possível, para abranger um grande número de futuros utilizadores dessa cartografia. Contudo, para utilizadores com necessidades de informação mais genérica, o mapa produzido na 4ª fase (69.6%), e até na 3ª (79.6%), apresenta uma exactidão temática boa a muito boa.

O último teste teve como objectivo fazer uma avaliação exploratória da potencialidade da análise orientada por objectos, pelo que se realizou uma classificação híbrida que combinou a delimitação de objectos derivados das imagens AWiFS (Figura 11) e a sua classificação na classe que ocupa maior área através da classificação ao nível do *pixel*. O mapa assim produzido apresentou maior valor de exactidão global (66.26%), comparativamente à mera classificação ao nível do *pixel* usada na classificação dos objectos (62.11% do teste 1 com o CMV). A melhoria da exactidão global está relacionada com a delimitação dos objectos, o que demonstra que a classificação ganha ao ter em conta uma análise de contexto. Isto quer dizer que existem relações entre *pixels* e padrões paisagísticos passíveis de serem interpretados nas imagens AWiFS e que são impossíveis de explorar com uma abordagem ao nível do *pixel*. Mais ainda, o uso de objectos na classificação híbrida permitiu eliminar à

partida o efeito de *salt and pepper*, isto é, o aspecto *salpicado* dos mapas derivados por classificação ao nível do *pixel*. Este aspecto deve-se à classificação de alguns *pixels* isolados em uma classe diferente da mancha onde os mesmos se inserem. Assim, este resultado demonstra que esta abordagem de classificação tem potencialidade para produzir cartografia mais exacta e deverá ser explorada com maior profundidade.

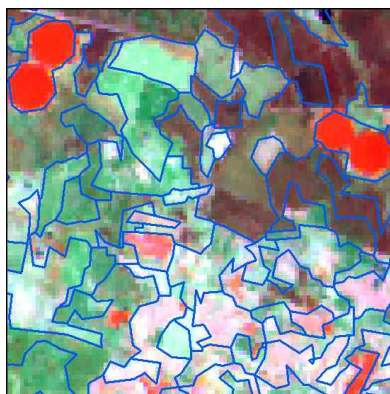


Figura 11 – Exemplo da segmentação das imagens AWiFS

No conjunto dos testes de classificação, nenhum atingiu o valor de 85% de exactidão global, que é usado regularmente como o valor mínimo de exactidão que a cartografia deve ter (Foody, 2002). O valor de 85% deriva do trabalho de Anderson *et al.* (1976), mas vários autores questionam a utilidade em usar o valor de 85% de exactidão como referência e rejeitam-no como critério para determinar o sucesso de um projecto de cartografia de ocupação do solo, pelo que não deve ser usado (e.g. Laba *et al.*, 2002; Wulder *et al.*, 2006). Além disto, Laba *et al.* (2002) referem que os projectos de cartografia de ocupação do solo a escalas regionais estabilizaram entre 50 a 70%, independentemente do detalhe temático ou da abordagem metodológica. Seguindo esta análise, pode-se dizer que os resultados obtidos pelos testes de classificação são satisfatórios.

No entanto, se comparados com os programas de ocupação do solo descritos do ponto 2.2, de uma maneira geral os resultados obtidos no presente trabalho ficam aquém, com excepção do Mapa de Ocupação do Solo da Grã-Bretanha. Este último caso apresenta uma menor qualidade, provavelmente por ser o programa que adoptou a metodologia mais simples, até na forma de avaliar a sua exactidão temática (sem um protocolo rigoroso de validação). Os restantes programas apresentam melhores resultados, ou assim o indiciam, como no caso do programa NLCD 2001, em que até ao momento foi apenas foi divulgada uma validação cruzada. A comparação mais pertinente de se fazer é com a cartografia de ocupação do solo produzida no programa indiano NR-Census (que usaram imagens AWiFS, árvores de decisão e obtiveram valores de exactidão global de 83% a 95%). A diferença nos resultados poderá estar relacionadas com várias questões. A diferença paisagística entre Portugal e Índia poderá ser um dos factores. Mesmo não considerando a estrutura da paisagem dos dois países, relacionada com o uso do solo, etc., de onde eventualmente

pode resultar numa estrutura mais fragmentada e difícil de classificar no caso português, é evidente que o regime climático da Índia, bastante diferente do mediterrânico, proporciona uma dinâmica e diversidade espectral relacionada com a fenologia da vegetação (nomeadamente, três épocas agrícolas). O aumento da diversidade e dinâmica espectral ao longo do ano é uma vantagem porque cria oportunidades para a distinção das classes. Associado a este factor, no projecto indiano foi usada uma grande quantidade de imagens AWiFS (imagens relativas a 10 meses do ano durante as três épocas agrícolas), o que permite possuir dados temporais suficientes para caracterizar todas as mudanças espectrais das classes dinâmicas. A conjugação de todos estes factores resulta na construção de condições mais favoráveis à produção de cartografia mais exacta, ainda que represente igualmente um acréscimo dos custos de produção.

#### **4.2. Variáveis dos testes de classificação**

A análise aos resultados dos testes de classificação permitem avaliar as seis variáveis referidas no ponto 3.4.3 e testadas ao longo do trabalho (classificador, classes de treino, tamanho da amostra de treino, variáveis de *input*, nomenclaturas de ocupação do solo e estratégias de classificação), bastante pertinentes no desenvolvimento de uma metodologia de classificação automática no âmbito dos objectivos do projecto COSMIC.

Relativamente à questão do classificador, os testes 1 a 12 mostram que o classificador da máxima verosimilhança (CMV) pode produzir cartografia de ocupação do solo tão ou mais exacta que as árvores de decisão (Tabela 16). Desta maneira, os resultados não corroboram os testemunhos encontrados na literatura especializada e os resultados de outros trabalhos de investigação (e.g. Friedl e Brodley, 1997; Pal e Mather, 2003). E no caso do CMV, o uso de probabilidades *a priori* não se revelou benéfica. Mas apesar dos resultados alcançados através de ambos os classificadores serem similares, o seu comportamento revelou-se bastante diferente.

A diferença entre os classificadores foi visível nas respostas às alterações das características da amostra de treino. Ao comparar os testes 1 e 3, 2 a 4, 9 e 11, 10 e 12 e ainda os testes 4 e 13, as árvores de decisão revelaram um comportamento similar perante a fase de treino com as classes de informação, com classes espectrais ou com o tamanho da amostra reduzida a metade. Ainda assim, revelaram um comportamento tendencialmente melhor com conjuntos de treino heterogéneos, isto é, não divididos em classes espectrais. O classificador da máxima verosimilhança demonstrou um comportamento inverso. Este mostrou-se mais exigente nas características da amostra de treino. Devido ao pressuposto que o CMV faz relativamente à normalidade dos dados, obriga a um trabalho de escrutínio das observações bastante exaustivo, o que leva à subdivisão das classes de ocupação do solo em várias classes espectrais. Mais ainda, comprovou-se que o CMV é sensível ao número de observações por classe. Isto porque exige que o número de observações por

classe de treino seja, no mínimo, 10 a 30 vezes o número de variáveis de *input* (Mather, 2004). Assim, para o CMV ser usado em condições ideais, é desejável conciliar a necessidade de subdividir as classes de ocupação do solo em classes espectrais normalmente distribuídas com a recolha de 10 a 30 vezes o número de variáveis de *input* para cada classe espectral. Este procedimento torna os custos da fase de treino mais elevados.

As variáveis de *input* não foram relevantes para a obtenção dos melhores resultados, como se pode observar na comparação dos testes 1 e 2 e nos testes 3 a 7. O uso de imagens de três datas não se mostrou vantajoso relativamente ao uso de apenas duas datas (Abril e Junho), assim como o NDVI. Este facto mostra que o mês de Outubro (pelo menos no que diz respeito à informação fornecida pelas imagens AWiFS) não acrescenta informação espectral útil para a discriminação das classes de ocupação do solo. O uso de apenas duas datas tem um impacte positivo na redução dos custos relativos à aquisição de imagens.

No que diz respeito às nomenclaturas de ocupação do solo, como seria de esperar, o uso de duas versões menos detalhadas nos testes 8 e 9 (com 10 e 6 classes) permitiu obter maiores valores de exactidão global relativamente ao teste 4 (com 15 classes). Os valores perto de 80% são bons, embora seja igualmente expectável e exigível que mapas com menor detalhe temático sejam mais exactos. Por este motivo, estes resultados demonstram que existem algumas confusões entre classes pouco semelhantes, o que impede de atingir valores de exactidão global ainda melhores. Por sua vez estes testes indiciam que a impossibilidade de alcançar valores de exactidão global mais elevados poderão estar relacionados com a capacidade das imagens AWiFS em fornecerem informação útil para classificar a paisagem portuguesa ao nível do *pixel*.

A adopção de duas estratégias de classificação por fases revelou-se pouco proveitosa. No caso da estratégia de classificar cada conjunto temporal de imagens de forma independente (teste 14), a exactidão global foi menor do que no teste homólogo (teste 4), e no caso da classificação por fases do teste 15, a exactidão global aumentou (igualmente em relação ao teste 4). Contudo, a estratégia do teste 15 é consumidora de tempo e a baixa melhoria alcançada não justifica a sua aplicação. Pelo contrário, a classificação híbrida produziu resultados promissores, apesar de ter sido testada com pouca profundidade. A melhoria alcançada na classificação através da integração dos objectos resultantes da segmentação das imagens AWiFS torna esta estratégia merecedora de mais desenvolvimento ou até mesmo testar a análise de imagens orientada por objecto.

Uma análise integral das seis variáveis testadas no conjunto dos testes de classificação faz das árvores de decisão o classificador mais indicado para um projecto como o COSMIC. Esta afirmação advém dos resultados entre os dois classificadores terem sido semelhantes na classificação de imagens AWiFS e das árvores de decisão serem muito menos exigentes

relativamente às características da amostra de treino. Numa classificação assistida, a fase de treino é a que se apresenta como a mais fulcral para a obtenção de bons resultados. Assim, ao usar um classificador que devolve resultados semelhantes independentemente das características da amostra de treino, assegura à partida uma fase de treino mais rápida e mais simples de se executar, assim como maior independência dos resultados em relação à influência subjectiva do analista que constrói a amostra de treino.

Estas questões ganham maior força num projecto que pretende produzir regularmente cartografia de ocupação do solo. A operacionalização do projecto será mais pesada no primeiro ano de produção, quando se tem que produzir a amostra de treino de raiz. Naturalmente, nos anos seguintes, o esforço será menor por via da reutilização da amostra de treino do ano anterior. Também nesta tarefa as árvores de decisão são vantajosas porque apenas é necessário identificar eventuais alterações de ocupação do solo de um ano para o outro, nas observações da amostra de treino. No caso do uso do CMV, eventualmente terá que se identificar alterações fenológicas e outras, mesmo em observações que mantêm a mesma ocupação do solo. Mesmo que no primeiro ano de operação do projecto fosse investido na construção de uma grande amostra de treino, que albergasse as observações suficientes para cada classe espectral (10 a 30 vezes o número de variáveis de *input* por classe espectral), no ano seguinte, a reutilização da amostra de treino continuaria a ser morosa. Isto porque algumas observações poderiam continuar a pertencer à mesma classe de ocupação do solo, mas mudar de classe espectral. Por exemplo, o caso de uma área agrícola de regadio passar a produzir mais que uma cultura por ano ou vice-versa. Por outro lado, a sensibilidade do CMV à relação do tamanho da amostra com a quantidade de variáveis de *input*, limita no futuro o uso de mais variáveis de *input*. Pelo contrário, a mesma amostra de treino criada para treinar as árvores de decisão para a classificação de três imagens AWiFS (12 bandas), pode ser usada com maior quantidade de imagens ou dados auxiliares, sem surgirem problemas relacionados com a dimensionalidade dos dados.

#### **4.3. Matrizes de confusão dos melhores testes ao nível do *pixel***

A exactidão global é um índice que fornece uma avaliação geral da qualidade da cartografia, mas pode ser insuficiente para determinados utilizadores. Estes podem ter interesse específico na exactidão de uma determinada classe de ocupação do solo ou na exactidão da área estimada (Foody, 2002). Nestes casos, a exactidão global é pouco clara, uma vez que cartografia com valores de exactidão global semelhantes podem apresentar resultados específicos por classes bastante distintos. Além disso, como referido no ponto 3.5, a amostragem estratificada origina diferentes intensidades de amostragem em função da área ocupada por cada classe, o que provoca algum enviesamento. Por estes motivos, para os testes de cada um dos classificadores que apresentaram maior exactidão global na Tabela

16 (não contando com o teste 16 que foi realizado apenas como uma exploração introdutória), e realizados com a nomenclatura de ocupação do solo padrão, são aqui analisadas as matrizes de confusão ponderadas com as áreas de ocupação de cada classe. Ou seja, são aqui analisadas as matrizes de confusão do teste 6, realizado com o CMV, e do teste 15, realizado com árvores de decisão. Recorde-se que cada observação da amostra de validação foi considerada bem classificada se um dos dois códigos (principal ou alternativo) correspondia com a classe produzida pelos classificadores. Nos casos em que nenhum dos códigos de referência concordava com a classe do mapa, é o código principal que consta nas matrizes.

Antes da análise às matrizes de confusão, é relevante notar que a exactidão global calculada a partir destas, é diferente da exactidão global que consta na Tabela 16. A diferença provém das diferentes fórmulas de cálculo. Na Tabela 16 constam os valores de exactidão global calculados com as fórmulas convencionais apresentadas na Figura 1, enquanto aqui, para uma análise mais aprofundada, foram calculados com a fórmula apresentada no ponto 3.5, com a ponderação da área ocupada por cada classe, como referido. A alteração da exactidão global está relacionada com a maior facilidade em classificar classes que ocupam menor área. Se uma determinada classe ocorre numa pequena percentagem de área classificada, irá apresentar menores diferenças espectrais intra-classe. Por exemplo, nas classes com vegetação, as diferenças espectrais intra-classe podem estar relacionadas com a fenologia, diferentes estados sanitários da vegetação, diferentes espécies, etc., enquanto nas classes artificiais podem estar relacionadas com a diversidade de materiais de construção. Logo, se estas classes ocorrem numa pequena área, os elementos que as constituem são mais semelhantes, os respectivos *pixels* são mais homogêneos e mais facilmente a amostra de treino representa a (pouca) variabilidade espectral das mesmas. Perante isto, com uma amostragem estratificada, onde é recolhido o mesmo número de observações para todos os estratos, o maior número de observações bem classificadas pertencentes às classes com pouca área irão favorecer o aumento do valor da exactidão global. Por este motivo é expectável que a exactidão global ponderada com a área ocupada por cada classe seja inferior. Naturalmente, juntamente com a exactidão global, os índices de qualidade específica das classes de ocupação do solo são também alterados.

De seguida são apresentadas e discutidas as matrizes de confusão ponderadas com a área ocupada por cada classe de ocupação do solo. No final, os índices de qualidade derivados da matriz de confusão do teste 6 com o CMV são comparados com os índices de qualidade derivados das matrizes construídas segundo a Figura 1 para observar o efeito da ponderação da área ocupada por cada classe (Tabela 20, página 68).

Relativamente ao mapa produzido no teste 6 com o CMV (Figura 12), para além da exactidão global (58.66%), a sua matriz de confusão (Tabela 17, página 62) demonstra que existem grandes diferenças de exactidão específica por classe.

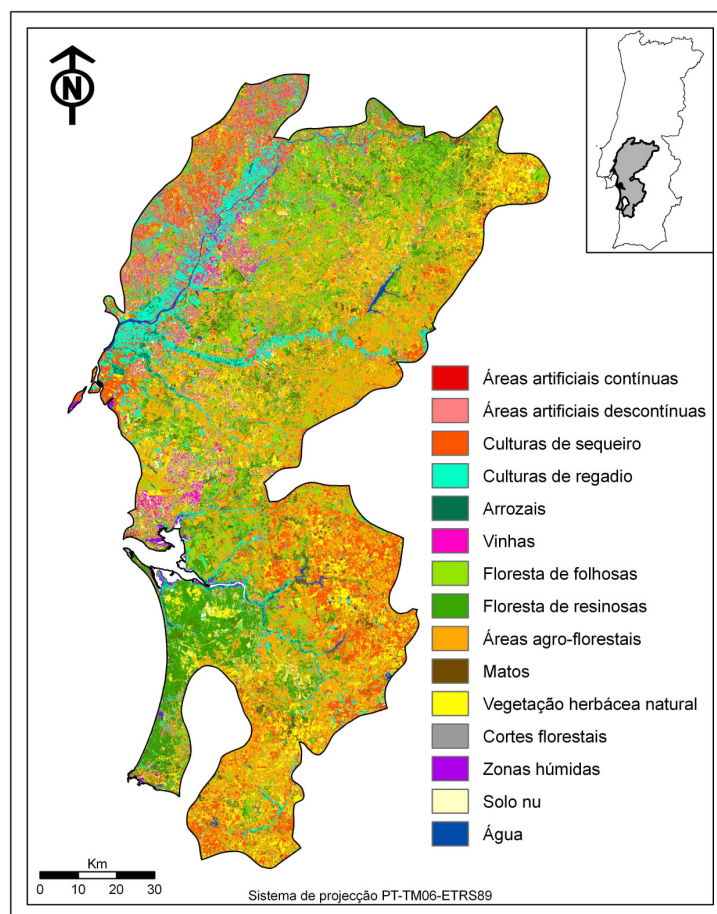


Figura 12 – Mapa de ocupação do solo produzido no teste 6 com o classificador da máxima verosimilhança

Em relação à exactidão do produtor (ou por outras palavras, erros de omissão), pode-se verificar que as classes 14 (Solo nu), 1 (Áreas artificiais contínuas), 12 (Cortes florestais), 11 (Vegetação herbácea natural), 8 (Floresta de resinosas) e 15 (Água) apresentam valores inferiores a 50%. As classes com menor erro de omissão foram as classes 3 (Culturas de sequeiro), 9 (Áreas agro-florestais), 13 (Zonas húmidas) e 4 (Culturas de regadio), com valores superiores a 70%.

No que diz respeito à exactidão do utilizador (erros de comissão), as classes pior representadas foram as classes 2 (Áreas artificiais descontínuas), 4 (Culturas de regadio) e 8 (Floresta de resinosas), com valores inferiores a 50%. Com mais de 70% apresentam-se as classes 15 (Água), 5 (Arrozais) e 7 (Floresta de folhosas).

Os baixos valores de exactidão do produtor (EP) e exactidão do utilizador (EU) de algumas classes indicam confusão. Parte dos erros é entre classes espectralmente próximas, por exemplo, entre florestas de resinosas e de folhosas, entre Floresta de resinosas e Matos e

entre Vegetação herbácea natural e Culturas de sequeiro ou Áreas agro-florestais. Estas três últimas classes (3, 9 e 11) constituem um trio onde as confusões são correntes. A confusão entre Culturas de sequeiro e a Vegetação herbácea natural pode ser explicada pelo tipo de vegetação ser muito semelhante e pela dependência directa que esta tem da precipitação, em ambas as classes. Por sua vez, a confusão destas com as Áreas agro-florestais pode ser explicada pelo seu subcoberto (com presença de herbáceas ou culturas de sequeiro), o que acaba por confundir classes claramente diferentes no ponto de vista paisagístico (devido à presença e ausência de vegetação arbórea), principalmente em Áreas agro-florestais mais abertas. A Figura 13 mostra uma área de montado correspondente a um *pixel* AWiFS (quadrado vermelho), em que é possível ver a grande percentagem de área exposta de sobcoberto correspondente a uma cultura de sequeiro, cujo comportamento espectral irá influenciar o sinal captado pelo sensor e criar confusão com a classe Culturas de sequeiro.



Figura 13 – Exemplo de possível confusão entre Áreas agro-florestais e Cultura de sequeiro presente no sobcoberto

A questão da influência que o sobcoberto das classes com vegetação arbórea parece ser também a causa dos baixos valores de exactidão da classe 8 (Floresta de resinosas). Esta classe existe sob a forma de povoamento bastante homogéneos e densos, até povoamentos mais abertos sem matos ou com matos. A grande variabilidade intra-classe, que levou a que fosse a classe que originasse mais classes espectrais (Tabela 9), torna mais difícil a sua amostragem e a consequente classificação.

A classe Área artificial contínua apresenta um grande erro de omissão, sendo classificada como Área artificial descontínua e solo nu. No primeiro caso, a confusão deriva dos elementos comuns que existem entre as duas classes de área artificial. No segundo caso, as classes são ambas sem vegetação, de uma forma geral, bastante reflectivas em todas as bandas das imagens AWiFS, e daí a confusão.

Também com um grande erro de omissão, aparece a classe 12 (Cortes florestais). À primeira vista, este valor poderá surpreender (porque apenas duas observações desta classe foram mal classificadas na amostra de validação), mas há que notar que estas observações mal classificadas correspondem a classes com uma grande área de ocupação,



o que provoca um elevado valor estimado de área real (Ng) da classe 12, relativamente à sua área classificada no mapa (Nh1). Apesar de menor, verificou-se algum erro de comissão desta classe. A confusão entre Cortes florestais e áreas não agrícolas com vegetação deveu-se, principalmente a operações de remoção de material vegetal (desmatações e limpeza em áreas florestais), o que provocou uma aproximação no perfil espectral destas classes e conduziu à interpretação de corte florestal por parte do classificador.

De estranhar foi a baixa exactidão do produtor da classe 15 (Água). Normalmente, esta é uma das classes simples de classificar, devido ao seu comportamento peculiar ao longo do espectro electromagnético. A causa para a baixa exactidão do produtor é a confusão com a classe 13 (Zonas húmidas). Através da análise visual do mapa, é possível ver que esta confusão ocorre no limite das massas de água interiores, onde, naturalmente, a profundidade da água é menor e onde se faz sentir a influência das oscilações de cota dessas massas de água. Por outro lado, no limite das massas de água ocorrem *pixels* de fronteira entre água e outras classes (e.g. classes com vegetação) de onde poderá resultar um sinal espectral semelhante à classe 13. À excepção dos *pixels* de fronteira, facilmente a classe 15 é classificada, como se pode ver através do maior valor de exactidão do utilizador (91.43%).

Também se observa confusão entre classes distintas, tais como, Floresta de folhosas e Áreas artificiais descontínuas, Solo nu e Vegetação herbácea natural ou ainda Vegetação herbácea natural e Áreas artificiais contínuas. A classe Áreas artificiais descontínuas é a que apresenta o menor valor de exactidão do utilizador (34.78%) enquanto a classe Solo nu é a que apresenta a menor exactidão do produtor (21.78%). No primeiro caso, pode ser argumentado que a heterogeneidade conceptual da classe (áreas urbanas intercaladas com outra qualquer área não urbana, desde culturas anuais, permanentes, etc.) faz com que estas áreas tenham um comportamento espectral semelhante às classes puras dos elementos seus constituintes, criando confusões com várias classes, como mostra a matriz. No segundo caso, embora exista também alguma heterogeneidade conceptual, uma vez que engloba áreas como praia, pedreiras, solo sem vegetação, etc., não era de esperar as confusões demonstradas pela matriz. Existem confusões com classes que parte do ano têm solo nu ou praticamente nu (e.g. nas culturas de sequeiro depois de colhidas e antes das vinhas ganharem folhagem), mas estas classes têm vegetação noutras épocas do ano, o que seria de esperar que levasse à sua distinção ao usar dados multitemporais. No caso da confusão entre Solo nu e Vegetação herbácea natural, à partida, classes muito diferentes, pode ser em parte explicada por duas situações. Em certos casos, a vegetação herbácea é quase esparsa devido às condições biofísicas do local. Noutros casos, antigas área florestais em recuperação depois de cortadas, a vegetação ainda não apresenta uma grande densidade, o que pode permitir que o solo interfira no sinal espectral registado pelo

sensor. Esta explicação é também, em parte, a causa para a confusão entre Vegetação herbácea natural e Áreas artificiais contínuas, que se verificou ocorrer em áreas de vegetação quase esparsa, em que o solo tem influência no sinal recebido pelo sensor. Esperava-se que as observações que demonstram confusão entre as classes 11 e 1 fossem classificadas ou na classe 11 ou na classe 14, pelo que lhes foi atribuído os dois códigos como principal e alternativo. No entanto, estas áreas foram classificadas com a classe 1, o que na verdade, demonstra a confusão entre as classes 1 e 14, já referida anteriormente.

Igualmente estranho é o baixo valor de exactidão do utilizador, isto é, o grande erro de comissão da classe 4 (Culturas de regadio). Devido à presença de vegetação verde no Verão, normalmente esta é uma classe facilmente classificada, apenas confundível com a classe 5 (Arrozais), devido ao ciclo cultural ser semelhante. Dois motivos podem explicar este resultado. Por um lado, as culturas de regadio são muitas vezes praticadas em pequenas parcelas, incluídas em sistemas parcelares complexos. Daqui resultam muitos *pixels* mistos, que são mais difíceis de classificar devido à influência espectral de várias classes. Por outro lado, devido à irrigação, é uma classe muito dinâmica (do ponto de vista da prática agrícola), o que provoca uma maior variabilidade temporal. Por outras palavras, à data da imagem AWiFS de Verão, existem culturas de regadio mais precoces que foram já colhidas, outras que se encontram com a vegetação verde no máximo do seu crescimento, e ainda, culturas no início do crescimento. Além do mais, existem parcelas onde se produz apenas uma cultura por ano (no Verão) ou mais (Verão, Primavera ou Outono). Esta variedade temporal e espectral traz problemas à classificação e causa erros.

As melhores classes que mantiveram um equilíbrio entre o erro de comissão e omissão foram as classes 3 (Culturas de sequeiro) e 7 (Floresta de folhosas) com a exactidão do produtor sempre acima de 62% e a exactidão do utilizador acima de 63%.

Tabela 17 – Matriz de confusão do teste 6 com o classificador da máxima verosimilhança

		Informação de referência															Total	Nh1	EU	P.Abs.
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15				
Mapa	1	20							1		1	1	5		2		29	6133.9	68.97	14.33
	2	3	8	1		1	1	4	1	1	1	2					23	13506.2	34.78	16.65
	3			39					1	3	2	10	1		4		61	142923.7	63.93	10.17
	4		1	1	17	3	2	6		1	2	5					38	34237.2	44.74	13.40
	5				3	30										1	34	18343.8	88.24	9.20
	6		2		1		12	2				2			2		21	26438.6	57.14	18.15
	7			1				61	14	3	2						81	132624.6	75.31	7.91
	8	2	1					7	23	1	7	3			3		47	47300.5	48.94	12.09
	9	1		5	1	1	3	5	1	30	1	9	1				58	147986.4	51.72	10.85
	10							4	5	4	19	2		1			35	61443.7	54.29	14.01
	11	1	1	3	1	1		3		4	2	28			5		45	57266.9	62.22	11.99
	12								3	4	1	2	18				28	4062.7	64.29	15.11
	13				1						1	1		14	8		22	10309.5	63.64	17.21
	14	3											1				13	10078.4	61.54	23.03
	15														3	32	35	4831.2	91.43	7.86
Total		30	13	50	25	35	18	93	50	46	40	69	20	18	24	39	570			
Ng		26070.1	24030.4	179600.6	47677.9	31363.8	46953.7	212206.4	104501.1	191236.2	104785.7	172045.3	12660.3	13996.4	46153.0	9861.0		EG	56.66	
P.Abs.		0.93	0.80	2.41	1.16	0.85	1.37	2.31	1.84	2.89	1.88	2.70	0.82	0.48	1.30	0.22		P.Abs.	3.989	
EP		23.53	56.21	79.58	71.81	58.49	56.31	62.50	45.26	77.38	58.64	33.29	32.09	73.66	21.84	48.99				
lc90%		10.81	21.01	8.80	16.15	19.31	19.13	7.65	10.35	7.90	11.24	7.37	25.82	28.06	9.57	13.62				

Ng - Estimativa da área real da classe (ha)

Nh1 - Estimativa da área da classe correctamente classificada

P.Abs - Precisão absoluta da estimativa para um nível de confiança de 90%

EU - Exactidão do utilizador

EP - Exactidão do rodutor

EG - Exactidão global

Ng - Estimativa da área real da classe (ha)  
 Nh1 - Estimativa da área da classe correctamente classificada  
 P. Abs - Precisão absoluta da estimativa para um nível de confiança de 90%  
 EU - Exactidão do utilizador  
 EP - Exactidão do rodutor  
 EG - Exactidão global

A matriz de confusão da Tabela 18 (página 66), relativa ao mapa produzido no teste 15 com árvores de decisão (Figura 14), com 57.73% de exactidão global, também demonstra confusão entre diversas classes, algumas diferentes do caso do teste 6 com o CMV, confirmando que mapas com uma exactidão global semelhante podem apresentar diferentes valores de exactidão por classe.

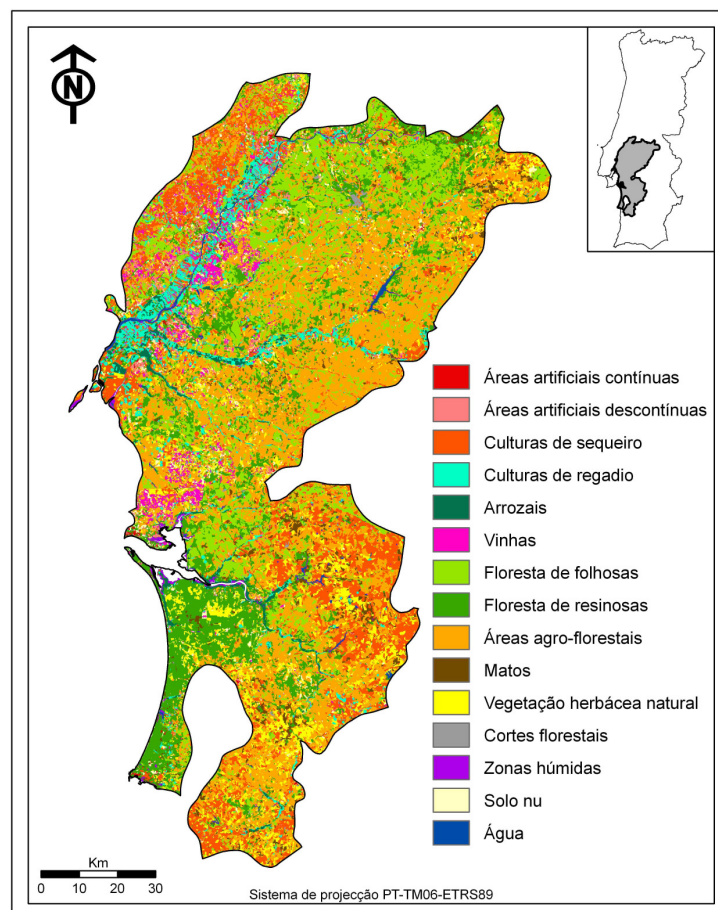


Figura 14 – Mapa de ocupação do solo produzido no teste 15 com árvores de decisão

Em relação à exactidão do produtor (ou por outras palavras, erros de omissão), pode-se verificar que as classes 1 (Áreas artificiais contínuas), 2 (Áreas artificiais descontínuas), 12 (Cortes florestais), 15 (Água), 14 (Solo nu), 10 (Matos), 11 (Vegetação herbácea natural), 6 (Vinhas) e 5 (Arrozais) apresentam valores inferiores a 50%. As classes com menor erro de omissão foram as classes 9 (Áreas agro-florestais), 3 (Culturas de sequeiro) e 8 (Floresta de resinosa), com valores superiores a 70%.

No que diz respeito à exactidão do utilizador (erros de comissão), as classes pior representadas foram as classes 2 (Áreas artificiais descontínuas), 6 (Vinhas), 8 (Floresta de resinosa), e 10 (Matos), com valores inferiores a 50%. Com mais de 70% apresentam-se as classes 1 (Áreas artificiais contínuas), 15 (Água), 5 (Arrozais), 7 (Floresta de folhosas) e 3 (Culturas de sequeiro).

Existem erros entre classes espectralmente próximas, nomeadamente, entre Floresta de folhosas e Floresta de resinosas, entre Floresta de resinosas e Matos e entre Áreas agro-florestais e Vegetação herbácea natural. A confusão entre estas classes é igualmente observada na matriz da Tabela 17, discutida acima e visível através do Anexo 6. No entanto, de referir que no teste 15, com árvores de decisão, as referidas confusões entre classes apresentam-se com ligeiras diferenças. Assim, a confusão entre as classes 8 (Floresta de resinosa) e 10 (Matos) surge num maior erro de comissão e omissão, respectivamente, o que acabou por penalizar mais a classe 10, visível na matriz da Tabela 18 com piores valores de EP e EU. Ainda em relação à classe 8, o teste 15 releva de forma mais acentuada a confusão entre esta classe e as classes 1 (Áreas artificiais contínuas) e 11 (Vegetação herbácea natural), que apesar de visíveis na matriz da Tabela 17, são menores. Nas áreas onde se verificou estas confusões, representadas por *pixels* mistos na amostra de validação, as árvores de decisão revelaram maior dificuldade para os classificar. A confusão entre as classes 8 e 1 ocorreu em *pixels* que continham algum material vegetal, nomeadamente algum espécime de resinosa ou folhosa (e.g. árvores na via pública de uma cidade) e em áreas artificiais menos reflectivas do que o habitual. A confusão entre as classes 8 e 11 ocorreu em *pixels* localizados em áreas heterogéneas, ocupados por vegetação herbácea e árvores dispersas, muitas vezes folhosas, que acabam por se confundir com espécies de resinosas.

Existem várias outras classes com um comportamento semelhante entre as duas matrizes. À semelhança do CMV, as árvores de decisão tiveram dificuldade em classificar a classe 1 (Áreas artificiais contínuas), a qual apresenta um grande erro de omissão, sendo classificada como Floresta de resinosas, como referido no parágrafo anterior, e Solo nu. A confusão entre as classes 1 e 14 (Solo nu) deriva de ambas as classes não apresentarem qualquer vegetação e serem bastante reflectivas em todas as bandas das imagens AWiFS, como referido anteriormente relativamente à matriz da Tabela 17. No teste 15, a classe 1 teve a particularidade de apresentar apenas erro de omissão. Relativamente à exactidão do utilizador, o valor de 100% indica que não existem erros de comissão. Por sua vez, a classe 14 (Solo nu) apresenta confusões entre Vegetação herbácea natural e Floresta de resinosas (à semelhança do CMV no teste 6), e ainda com Áreas agro-florestais (neste caso apenas visível na matriz de confusão da Tabela 18). A confusão entre Solo nu e Vegetação herbácea natural pode estar relacionada com condições biofísicas e com a fase de recuperação e antigos povoamentos florestais, como referido. No segundo e terceiro caso, a confusão existe em áreas ocupadas por sobreiros, pinheiros mansos e instalados em solos arenosos, muitas vezes expostos. Assim, nestas áreas existe componentes comuns entre três classes e as árvores de decisão foram mais favoráveis à classificação destes *pixels* na classe 9 (Áreas agro-florestais) do que o CMV.

A classe 12 (Cortes florestais), é uma classe com um comportamento em tudo semelhante com os dois classificadores relativamente à EP e EU, visível em ambas as matrizes e discutido acima. O mesmo se pode dizer da classe 15 (Água) que apresenta índices de qualidade específicos semelhantes nas duas matrizes (baixo valor de EP e elevado valor de EU). A causa para a baixa exactidão do produtor é a confusão com a classe 13 (Zonas húmidas) no limite das massas de água interiores.

Em relação à classe 2 (Áreas artificiais descontínuas) verifica-se a dificuldade na sua classificação, também com árvores de decisão, e é uma das classes com piores valores de EP e EU. A heterogeneidade conceptual desta classe referida acima parece ser assim determinante para o comportamento deficiente de ambos os classificadores na sua classificação.

A classe 6 (Vinhas) não foi bem classificada por ambos os classificadores nos testes 6 e 15. No caso do teste 15, os valores de exactidão são piores. À partida, era esperado que os classificadores sentissem dificuldade em classificar a classe 6 devido à própria estrutura agrícola caracterizada pela pouca densidade de folhagem, que resulta na influência do solo no sinal registado pelo sensor, e tamanho diminuto e distribuição em mosaico de algumas vinhas. Portanto, não é de estranhar a confusão com classes como Solo nu ou Vegetação herbácea natural.

Apesar de o resultado obtido com árvores de decisão do teste 15 apresentar um valor de EG ligeiramente inferior ao resultado obtido com o CMV do teste 6 e um decréscimo na exactidão de algumas classes, verificou-se a melhoria de outras. A classe 8 (Floresta de resinosas) foi melhor classificada no teste 15, assim como a classe 3 (Culturas de sequeiro). Relativamente às melhores classes que mantiveram um equilíbrio entre o erro de comissão e omissão, também no teste 15 foram as classes 3 (Culturas de sequeiro) e 7 (Floresta de folhosas) com a exactidão do produtor acima de 63% e a exactidão do utilizador acima de 70%.

Tabela 18 – Matriz de confusão do teste 15 com árvores de decisão

Mapa

Informação de referência															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	12														12
2	2	2	2	1		1			1		1				10
3			29	2		1	2	1		1	3	1		1	41
4				14	3		4								22
5				2	26										28
6		2		4		7			1	1	3				18
7			1				63	10	2	2	1				79
8	6	2		1	3		13	38	5	10	10			3	92
9			3		1	3	7		46	4	20	2		4	90
10			1					2	2	10	5		2		21
11	2	2	5	1		1	1	2	1	4	33			3	55
12									4	1	1	16			22
13								1		1			15		24
14	5	1				2			1	1	1			14	25
15													2		31
Total	27	9	41	25	33	15	90	52	63	35	78	19	19	25	570
Ng	29209.1	16663.2	136877.6	54659.8	32331.5	33611.8	215139.2	111785.2	208509.7	85831.1	205414.8	16591.2	13452.3	47269.1	15796.0
P.Abs.	0.89	0.72	1.88	1.17	0.88	1.20	2.17	1.73	2.68	1.83	2.75	0.88	0.54	1.37	0.58
EP	10.75	21.56	77.33	60.77	46.13	36.15	65.63	72.45	86.14	33.36	34.12	30.93	54.87	32.23	31.55
Ic90%	4.01	20.40	8.90	14.02	15.46	18.24	7.33	9.53	5.53	11.29	6.59	20.52	26.60	12.72	14.33
Ng - Estimativa da área real da classe (ha) Nh1 - Estimativa da área da classe correctamente classificada P.Abs - Precisão absoluta da estimativa para um nível de confiança de 90% EU - Exactidão do utilizador EP - Exactidão do rodutor EG - Exactidão global															
															EG
															P.Abs.
															57.73
															3.745

Para facilitar a interpretação das matrizes de confusão, no Anexo 6 pode-se consultar a informação das mesmas em formato gráfico. Foram construídos gráficos de barras para cada uma das 15 classes de ocupação do solo dos melhores testes de forma a permitir uma

comparação visual entre os resultados dos dois classificadores. Em cada um destes gráficos é possível observar o erro de omissão e comissão de cada classe, assim como a exactidão do produtor e do utilizador. A análise adicional e conjunta dos gráficos permite tecer e confirmar considerações já referidas acima acerca das classes e classificadores (Tabela 19).

Classe de ocupação do solo	Breves considerações acerca das classes e classificadores
1 (Áreas artificiais contínuas)	Classificada de forma diferente entre os dois classificadores. As árvores de decisão não produziram erros de comissão, apenas e mais erros de omissão do que o CMV
2 (Áreas artificiais descontínuas)	Classe onde ocorreram os maiores erros em ambos os classificadores (com maior gravidade no caso das árvores de decisão)
3 (Culturas de sequeiro) e 4 (Culturas de regadio)	Classes com comportamentos semelhantes em ambos os classificadores, com uma ligeira vantagem para as árvores de decisão
6 (Vinhais)	Classificada com um maior equilíbrio entre erros de omissão e comissão com o CMV, inferiores a 50%
7 (Floresta de folhosas) e 8 (Floresta de resinosas)	Classes classificadas de forma semelhante por ambos os classificadores mas de forma um pouco mais exacta por parte das árvores de decisão
9 (Áreas agro-florestais)	Classificadas com menor erro de omissão com árvores de decisão
10 (Matos)	Apresenta erros de omissão e comissão inferiores a 50% na classificação com o CMV e superiores a 50% pelas árvores de decisão
11 (Áreas agro-florestais)	Apresenta valores de exactidão do utilizador e do produtor semelhantes com ambos os classificadores, mas com algumas diferenças na distribuição do erro de omissão nas classes 3 e 9
12 (Cortes florestais)	Classe com um comportamento muito semelhante com os dois classificadores
13 (Zonas húmidas)	Classificada com um erro de omissão ligeiramente superior no caso das árvores de decisão
14 (Solo nu)	Foi uma das classes onde os classificadores mostraram maiores dificuldades e existiu maior erro de omissão com o CMV e maior erro de comissão com as árvores de decisão
15 (Água)	Classe melhor classificada com o CMV, onde a maioria dos erros se registaram com a classe 13, enquanto as árvores de decisão produziram confusões com maior número de classes

Tabela 19 – Breves considerações acerca das classes e classificadores com base nas matrizes de confusão e nos gráficos do Anexo 6

Para finalizar a discussão sobre as matrizes de confusão, na Tabela 20 estão resumidas as diferenças de valores dos índices de qualidade (exactidão global, do utilizador e do produtor) quando calculados sem e com a ponderação da área ocupada por cada classe, relativamente ao mapa do teste 6 com o CMV.



Classe	Não ponderada EG=62.98%		Ponderada EG=58.66%	
	EU	EP	EU	EP
1	68.97	66.67	68.97	23.53
2	34.78	61.54	34.78	56.21
3	63.93	78.00	63.93	79.58
4	44.74	68.00	44.74	71.81
5	88.24	85.71	88.24	58.49
6	57.14	66.67	57.14	56.31
7	75.31	65.59	75.31	62.50
8	48.94	50.00	48.94	45.26
9	51.72	60.00	51.72	77.38
10	54.29	47.50	54.29	58.64
11	62.22	40.58	62.22	33.29
12	64.29	90.00	64.29	32.09
13	63.64	77.78	63.64	73.66
14	61.54	33.33	61.54	21.84
15	91.43	82.05	91.43	48.99

Tabela 20 – Valores de exactidão global (EG) não ponderada e ponderada do teste 6 com CMV

Ao observar-se a Tabela 20, verifica-se que alguns índices de qualidade variam bastante. Como referido, esta variação está relacionada com a maior facilidade em classificar classes que ocupam menor área. Assim, os erros de classificação observados em classes que ocupam uma grande área têm um maior impacto nos cálculos dos índices de qualidade, o que provoca a sua diminuição. Pelo contrário, os erros de classificação observados em classes que ocupam uma pequena área provocam a subida dos índices de qualidade. Mas como a ocorrência de erros nas classes que ocupam pouca área são menos frequentes, a exactidão global tende a baixar. Embora a matriz ponderada com a área ocupada por cada classe de ocupação do solo evite o enviesamento causado pela diferente intensidade de amostragem nas diferentes classes (decorrente da amostragem aleatória estratificada), a sua aplicação não é corrente. A situação mais comum é a derivação dos índices de qualidade segundo a Figura 1, o que conduz a uma estimativa optimista da qualidade da cartografia.

#### 4.4. Comparação dos melhores testes ao nível do *pixel* com o CLC2000

Outra forma de avaliar a cartografia produzida é a sua comparação com outro produto cartográfico. Para este efeito, os mapas produzidos pelos melhores testes de cada classificador, ao nível do *pixel*, foram comparados com o mapa CLC2000. Este é um mapa de referência e de grande utilização, com uma exactidão global de 82,8% (Caetano *et al.*, 2006), unidade mínima cartográfica de 25 ha e produzida por interpretação visual de imagens de satélite. Devido à existência de uma grande diferença da unidade mínima cartográfica do mapa CLC2000 e dos mapas produzidos nos testes de classificação, os últimos foram sujeitos a um filtro de maioria numa janela de 5x5 *pixels* para, assim, atenuar o efeito *salt and pepper* e generalizar pequenas manchas. Do resultado desta operação, apenas as áreas superiores a 25 ha foram usadas para comparar com o mapa CLC2000. Desta operação resultou que o mapa do teste 6 com o CMV e o mapa do teste 15 com árvores de decisão apresentaram, respectivamente, 80.64% e 80.65% da sua área com mais de 25 ha e foram estas áreas as usadas na comparação com o mapa CLC2000. Outra consequência das diferentes unidades mínimas cartográficas é as diferentes nomenclaturas de ocupação do solo e a necessidade de as compatibilizar. Assim, admitiu-se que classes heterogêneas do mapa CLC2000 correspondem a mais do que uma classe dos mapas dos testes de classificação. Por exemplo, considerou-se concordante a classe 3.1.1 (Florestas mistas) com as classes 7 (Floresta de folhosas) ou 8 (Floresta de resinosas). Nas matrizes de confusão da Tabela 21 e Tabela 22, pode-se observar a percentagem de área comparada concordante (a sublinhado) e discordante entre as classes de cada mapa. No Anexo 3 pode-se consultar a nomenclatura do mapa CLC2000.

		CLC2000																															
		1.1.1																				3.2.2				3.3.1							
		1.2.x																				3.2.3	3.2.1	4.x.x	3.3.2	5.x.x	222	223	241	242	243	313	324
		1.3.x	1.1.2	2.1.1	2.1.2	2.1.3	2.2.1	3.1.1	2.4.4	3.1.2	3.2.3	3.2.1	4.x.x	3.3.2	5.x.x	222	223	241	242	243	313	324											
Mapa do teste 6 com o CMV	1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0											
	2	0.1	0.4	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.4	0.1	0.0	0.0											
	3	0.0	0.0	7.8	2.3	0.3	0.1	2.1	4.5	0.0	0.1	0.2	0.0	0.0	0.0	0.2	1.6	0.6	1.7	0.6	0.1	0.3											
	4	0.0	0.0	0.2	3.0	1.1	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.1	0.1	0.8	0.1	0.1	0.1											
	5	0.0	0.0	0.0	0.6	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0											
	6	0.0	0.0	0.0	0.1	0.0	1.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.7	0.1	0.0	0.0											
	7	0.0	0.0	0.2	0.1	0.1	0.0	10.0	0.3	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.8	2.8	1.6											
	8	0.1	0.1	0.0	0.0	0.0	0.0	1.2	0.1	3.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.9	0.8											
	9	0.0	0.0	1.0	0.4	0.3	0.3	16.5	4.9	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.8	0.2	1.0	0.9	2.4	0.9											
	10	0.0	0.0	0.3	0.0	0.0	0.0	3.6	0.9	0.5	0.2	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.1	0.2	0.9	0.9											
	11	0.0	0.0	1.0	0.1	0.0	0.0	1.0	0.8	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.2	0.1	0.2	0.3											
	12	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1											
	13	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0											
	14	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1											
	15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0											
		49.2% de concordância																															

49.2% de concordância

Tabela 21 – Comparação entre o mapa do teste 6 com o CMV e o mapa CLC2000



Portugal Continental. Neste sentido, Costa *et al.* (2008) compararam as imagens MERIS usadas no trabalho de Carrão *et al.* (no prelo) com as imagens AWiFS. As imagens dos dois sensores foram usadas para derivar cartografia de ocupação do solo em condições semelhantes (relativamente à nomenclatura, à informação espectral e à informação temporal de cada sensor) e em condições diferentes (com recurso a toda a informação espectral e temporal disponível de cada sensor). Com informação espectral equivalente entre as duas imagens, a exactidão das classificações começou por ser semelhante e aumentou em ambos os casos à medida que se adicionou informação temporal. Ou seja, os resultados obtidos neste estudo foram semelhantes entre as classificações das imagens MERIS e AWiFS com condições equivalentes. Os resultados divergiram quando se adicionou toda a informação temporal (12 meses) e/ou espectral (13 bandas) disponível nas imagens MERIS. Logo, os resultados demonstram que os menores valores de exactidão obtidos na classificação de imagens AWiFS podem ser explicados pela sua menor resolução espectral (Tabela 2) e à menor informação temporal (Tabela 3).

Relativamente à informação temporal, o uso de apenas três datas deveu-se a um compromisso entre os custos de aquisição das imagens e a informação suficiente para classificar a ocupação do solo. As imagens de Primavera, Verão e Outono foram seleccionadas com o objectivo de adquirir as maiores diferenças espectrais que as classes apresentam ao longo do ano, mas não existe esta garantia. Principalmente em relação a classes mais dinâmicas como as classes agrícolas, que podem apresentar diferentes estados espectrais no mesmo período de tempo devido a diferentes práticas agrícolas (e.g. data de sementeira e colheita). Assim, naturalmente que a disponibilidade de mais imagens de diferentes alturas do ano contribuiria para o aumento da exactidão da classificação, como no caso do trabalho de Carrão *et al.* (no prelo) e corroborado por Costa *et al.* (2008). Esta conclusão não contradiz os resultados do presente trabalho, que como já foi referido, mostrou que a imagem AWiFS de Outubro não exerceu papel relevante da melhoria dos resultados. Isto apenas quer dizer que das três datas usadas no presente trabalho, as imagens de Abril e Julho foram suficientes, o que não significa que o uso de mais imagens de outras alturas do ano não seja benéfico.

Relativamente à informação espectral, o estudo de Costa *et al.* (2008) demonstrou que a maior resolução espectral das imagens MERIS permite a obtenção de melhores resultados do que as imagens AWiFS. Embora o sensor AWiFS registe radiação electromagnética na mesma região que outros sensores usados com sucesso na classificação da ocupação do solo de Portugal Continental, apenas tem quatro bandas com uma considerável largura de banda. Tendo em conta a região do espectro comum a ambos, enquanto o sensor AWiFS tem três bandas divididas pela radiação verde, vermelho e infra-vermelho próximo, o sensor MERIS tem várias bandas (mais estreitas) em cada uma destas regiões (Figura 15). Assim, duas classes de ocupação do solo que tenham um comportamento diferente numa

determinada região do espectro, eventualmente podem ser distinguidas através das diferenças registadas pelas várias bandas dessa região do sensor MERIS, enquanto a única banda dessa região do sensor AWiFS registará um valor semelhante para ambas as classes. Estas diferenças na forma como os sensores registam a radiação captada pode ser bastante influente no processo de classificação de imagens uma vez que a resolução espectral é a característica técnica das imagens de satélite mais influente na classificação da paisagem portuguesa (Carrão *et al.*, 2008; Costa *et al.*, 2008).

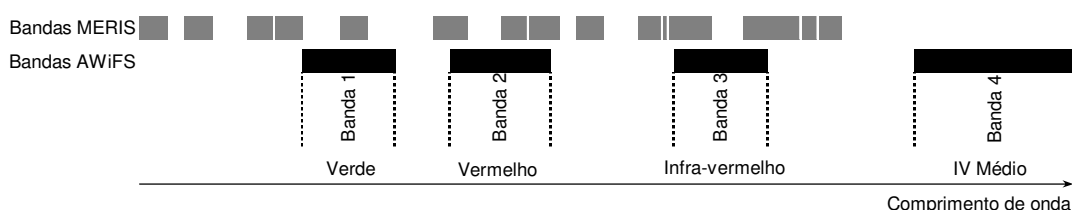


Figura 15 – Largura das bandas dos sensores AWiFS e MERIS

De notar que a quarta banda do sensor AWiFS (inexistente no sensor MERIS) é uma mais valia porque adquire informação na região do infra-vermelho médio (bastante útil para distinguir classes com vegetação). De facto, esta banda contribuiu muito para o poder discriminativo do sensor AWiFS, relativamente às bandas 1, 2 e 3 (Costa *et al.*, 2008). Todavia, foi insuficiente para derivar cartografia com valores de exactidão global mais elevados.

#### 4.6. Críticas e perspectivas futuras

O trabalho apresentado seguiu uma linha de desenvolvimento que se considerou adequada dentro dos objectivos definidos, mas qualquer projecto de classificação de imagens de satélite inclui erros que, naturalmente, se querem reduzidos, e para tal, não devem ser ignorados na definição e execução da metodologia. Assim, parte dos resultados obtidos poderão estar relacionados com a influência conjunta ou individual de erros de amostragem, erros no registo das imagens e erros na informação de referência. Contudo, estes tipos de erros estão sempre presentes e esta consciência leva a que e a sua ocorrência e magnitude sejam mínimas e não tenham expressão nas conclusões que os resultados permitem construir.

Relativamente à avaliação dos resultados, à que ter em conta que foi realizada com base numa amostra de validação de tamanho reduzido, como referido no ponto 3.5. Desta forma, deve-se ter em mente que os resultados obtidos têm uma precisão absoluta considerável (principalmente em relação à exactidão do utilizador e do produtor). Assim, a amostra de validação usada, suficiente para desenvolver os testes de classificação, deverá ser melhorada de forma a se obter estimativas mais exactas acerca da qualidade da cartografia produzida.

Outro aspecto importante reside na interpretação das matrizes porque, aquando da sua construção, é excluído o código alternativo das observações da amostra de validação. Assim, nas matrizes de confusão, podem surgir erros entre classes espectralmente afastadas, que na realidade ocorrem em observações que correspondem a *pixels* mistos, nos quais o código alternativo (excluído na construção das matrizes) é mais semelhante com a classe do mapa do que o código principal.

Devido aos custos inerentes à aquisição de imagens, não é possível dispor de uma grande quantidade de dados multitemporais AWiFS como no caso do projecto ao abrigo do programa indiano NR-Census (ver ponto 2.2.5). Caso o fosse, talvez os resultados fossem superiores. Ainda assim, existe espaço para averiguar se é possível melhorar os resultados através do uso de outras imagens AWiFS. Os resultados demonstraram que a imagem de Outubro não teve um papel relevante da exactidão da classificação. Por outro lado, os resultados também demonstraram dificuldades em classificar algumas áreas de classes mais dinâmicas, como é o caso das classes agrícolas. Esta dificuldade está relacionada com o diferente estado espectral de áreas que pertencem à mesma classe de ocupação do solo no momento a que as imagens dizem respeito. Já foi referido o exemplo das culturas de regadio, que à data da imagem AWiFS de Verão, existem culturas mais precoces que foram já colhidas, outras que se encontram com a vegetação verde no máximo do seu crescimento, e ainda, culturas no início do crescimento. Assim, poderá ser proveitoso substituir a imagem de Outono e usar duas imagens de Verão, adquiridas no início e fim da estação.

Relativamente à nomenclatura adoptada, a sua definição consistiu na adaptação de uma versão menos detalhada da nomenclatura LANDEO com base nos resultados que se foram obtendo. De todas as 15 classes definidas, a que foi classificada com maiores erros foi a classe 2 (Áreas artificiais descontínuas), pelo que a sua exclusão da nomenclatura pode ser aconselhável. A decisão de excluir a classe 2, como o foi para outras classes da nomenclatura LANDEO original, tem a contrariedade de produzir uma nomenclatura menos detalhada, o que pode não interessar a parte dos potenciais futuros utilizadores. A solução para este problema poderá ser a classificação híbrida ou orientada por objectos. Nestas abordagens, as áreas artificiais descontínuas poderão ser identificadas através de análise de contexto ao nível do objecto.

A classificação híbrida adoptada no último teste mostrou que a segmentação das imagens AWiFS e a integração de objectos na classificação são uma linha de desenvolvimento a aprofundar. O bom resultado obtido no teste 16 deixa em aberto a possibilidade de usar a classificação híbrida com a nomenclatura de 15 classes na classificação dos objectos. Ou ainda, usar uma nomenclatura ao nível do *pixel* menos detalhada (e.g. sem Áreas artificiais descontínuas) e posteriormente classificar os objectos numa outra nomenclatura mais detalhada com base numa análise de contexto (e.g. com Áreas artificiais descontínuas,

Mosaicos). Além do maior valor de exactidão global obtido, esta abordagem permite a derivação de mapas mais próximos da preferência dos seus utilizadores finais. Ou seja, permite a derivação directa de mapas em formato vectorial e sem o aspecto *salt and pepper*, o que se torna mais atractivo em termos visuais, como se pode observar através da comparação do mapa derivado no teste 1 com CMV e a sua combinação com os objectos no teste 16 (Figura 16).

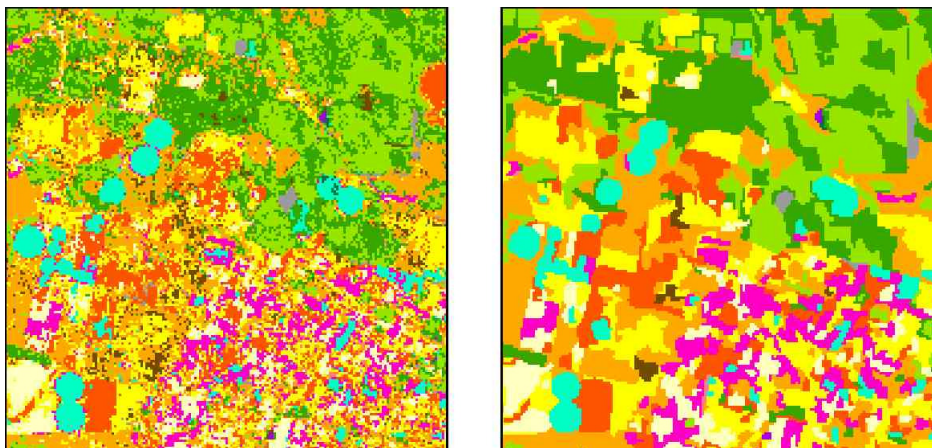


Figura 16 – Comparação entre o aspecto visual das classificações ao nível do *pixel* e híbrida *pixel/objecto*

Tanto no caso da classificação ao nível do *pixel* como na classificação híbrida, a adopção das árvores de decisão como classificador é vantajosa. Esta vantagem advém aquando da operacionalização do projecto e da extensão da metodologia a toda a área continental portuguesa. Como afirmam os resultados, a adopção deste classificador poderá abreviar o tempo de construção das amostras de treino dos restantes estratos não incluídos da área de estudo. Posteriormente, o uso de árvores de decisão deixa sempre em aberto a possibilidade de usar mais variáveis de *input* com a mesma amostra de treino, quer sejam mais imagens, índices de vegetação ou informação auxiliar integrada com as imagens na fase de classificação.

Fica ainda por avaliar os resultados obtidos em áreas cobertas por nuvens e suas sombras. Uma vez que na área de estudo (estrato 214) não existem nuvens em qualquer uma das imagens, não foi adoptada nenhuma metodologia alternativa. Futuramente, na classificação de outros estratos, a máscara de exclusão destas áreas, construída no pré-processamento das imagens, terá como função definir as áreas onde apenas se usará as datas que não apresentam nuvens. No caso de áreas cobertas por nuvens apenas em Outubro, os resultados deverão ser semelhantes aos apresentados pelos testes de classificação. Nos outros casos, é expectável que a exactidão seja inferior. De qualquer maneira, a área afectada por nuvens é reduzida, pelo que não deverá ter um grande impacte da exactidão global na classificação da totalidade do território português.

## 5. Conclusões

De uma forma geral, a exactidão global da cartografia produzida com base nas imagens AWiFS foi aceitável, dentro da magnitude de resultados apresentados por outros projectos (Laba *et al.*, 2002), independentemente das várias abordagens metodológicas testadas. Os resultados obtidos com o CMV e as árvores de decisão foram similares, mas apesar disso, as respostas de cada um às características da amostra de treino foram bastante diferentes. Esta diferença deve-se ao facto do CMV requerer um número mínimo de observações por classe de treino e que estas tenham uma distribuição normal. A amostra de treino usada nos testes não cumpre integralmente estes requisitos porque o elevado tempo que seria necessário para recolher um tal número de observações tornaria a metodologia muito morosa.

As árvores de decisão foram insensíveis às características da amostra de treino. Os melhores resultados foram produzidos com amostras de treino heterogéneas, ou seja, com o uso das classes de informação como classes de treino, em vez de classes espectrais. Isto significa que as árvores de decisão não exigem que a amostra se caracterize pela distribuição normal. Este facto torna desnecessária a divisão das classes de ocupação do solo em classes espectrais. Por outro lado, não são exigentes relativamente ao número de observações por classe. Estes dois aspectos permitem que o desenvolvimento da fase de treino seja bastante mais rápida e sem perdas na exactidão da classificação. Assim, as árvores de decisão são um classificador vantajoso no âmbito de um programa operacional e anual de produção de cartografia de ocupação do solo, nomeadamente no âmbito do projecto COSMIC.

Para obter os melhores resultados, as imagens AWiFS de Julho e Abril foram suficientes, pelo que a aquisição destas imagens no mês de Outubro não aumenta a informação útil para classificar a paisagem portuguesa. Assim, o uso de apenas duas datas (Abril e Julho) é um aspecto positivo para a redução de custos.

A adopção de estratégias de classificação por fases não se revelou benéfica para melhorar os resultados relativamente ao aumento de complexidade na sua execução. Por outro lado, a segmentação das imagens AWiFS e a integração de objectos numa classificação híbrida juntamente com a classificação ao nível do *pixel* mostrou ser um caminho a desenvolver.

Segundo Chuvieco (1996), a exactidão da cartografia depende, principalmente, da metodologia adoptada e da qualidade dos dados. Partindo deste princípio, a insistência das várias abordagens de classificação testadas em derivar cartografia com valores modestos de exactidão temática, indicia que é a informação contida nas imagens AWiFS uma das principais causas dos resultados modestos e não a ineficiência das abordagens metodológicas ou as características da amostra de treino. Deste modo, e voltando à questão do desempenho dos dois classificadores testados, a baixa resolução espectral das



imagens AWiFS, o que representa uma limitação, como constatado por Costa *et al.* (2008), poderá ser o motivo pelo qual as árvores de decisão não terem superado o CMV, no que diz respeito à exactidão global. Perante uma resolução espectral limitada para classificar a paisagem portuguesa, ambos os classificadores apresentaram capacidades suficientes para classificar as imagens. Eventualmente, perante imagens que fornecessem mais informação, as árvores de decisão poderiam demonstrar maiores capacidades para as classificar, e assim, corroborar os trabalhos que as relatam como mais exactos do que o classificador da máxima verosimilhança.

Considerando todos os resultados obtidos, dentro da aproximação metodológica desenhada, considero que a melhor abordagem passa por usar as árvores de decisão para classificar imagens AWiFS de Abril e Julho. Como referido, as árvores de decisão são vantajosas no que respeita ao desenvolvimento da fase de treino e o uso da imagem de Outubro não se traduziu num acréscimo significativo da exactidão temática. Relativamente à nomenclatura de ocupação do solo, é de considerar várias hipóteses, talvez mediante a produção de mais do que um mapa. Além da nomenclatura de 15 classes, o uso de 10 classes (como no teste 10) pode originar um produto bastante atractivo para utilizadores menos exigentes em detalhe temático.

Apesar das conclusões apresentadas, o trabalho de investigação e desenvolvimento de uma abordagem metodológica eficiente e executável no âmbito dos objectivos do projecto COSMIC não está esgotado. O resultado obtido no teste 16 indica que as imagens AWiFS incorporam informação relacionada com padrões espaciais cuja exploração poderá superar a abordagem ao nível do *pixel*. Desta forma, deverão ser realizados mais testes em que a abordagem de classificação seja substituída ou associada à análise orientada por objectos.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Anderson, J. R., Hardy, E., Roach, J., e Witmer, R., 1976. *A land use and land cover classification system for use with remote sensor data*. Professional Paper 964, U.S. Geological Survey, Washington.
- Araújo, A. e Caetano, M., 2006. *Nomenclatura de ocupação do solo LANDEO*. Lisboa: Instituto Geográfico Português.
- Bauer, E., Burk, T., Ek, A., Coppin, P., Lime, S., Walsh, T., Walters, D., Befort, W. e Heinzen, D., 1994. Satellite inventory of Minnesota forest resources, *Photogrammetric Engineering & Remote Sensing*, 60, 287-298.
- Bossard, M., Feranec, J. e Otahel, J., 2000. *CORINE land cover technical guide-Addendum 2000*. Technical report n. 40, European Environment Agency, Copenhagen.
- BSR, 2006. *Guidelines for land use mapping in Australia: principles, procedures and definitions*, 3rd edition, Canberra: Australian Government, Bureau of Rural Sciences.
- Caetano, M., Santos, T., e Gonçalves, L., 2002. Cartografia de Ocupação do Solo com Imagens de Satélite: Estado da Arte, [CD-ROM]. In *Actas do VII Encontro de Utilizadores de Informação Geográfica (ESIG 2002)*, Oeiras, Portugal.
- Caetano, M., Mata, F., e Freire, S., 2006. Accuracy assessment of the Portuguese CORINE Land Cover map. In *Global Developments in Environmental Earth Observation from Space*, editado por A. Marçal. Rotterdam: Millpress, 459-467.
- Caetano, M., Pereira, M., Carrão, H., Araújo, A., Nunes, A., e Nunes, V., 2008. Cartografia temática de ocupação/uso do solo do Instituto Geográfico Português. *Mapping - Revista Internacional de Ciencias de la Tierra*, 126, 78-87.
- Campbell, J., 1981. Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogrammetric Engineering & Remote Sensing*, 47, 355-363.
- Card, D., 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy, *Photogrammetric Engineering & Remote Sensing*, 48, 431-439.
- Carrão, H. e Caetano, M., 2006. *Land Cover Cartography Accuracy Assessment: An approach in the framework of LANDEO project*. Lisboa: Instituto Geográfico Português.
- Carrão, H., Caetano, M. e Coelho, P., 2007. Sample Design and Analysis for Thematic Map Accuracy Assessment: An Approach Based on Domain Estimation for the Validation of Land Cover Products, [CD-ROM]. In *Proceedings of the 32nd International Symposium on Remote Sensing of Environment*, San Jose, Costa Rica.
- Carrão, H., Gonçalves, P. e Caetano, M., 2008. Contribution of multispectral and multitemporal information from MODIS images to land cover classification. *Remote Sensing of Environment*, 112, 986-997.
- Carrão, H., Gonçalves, P. e Caetano, M., no prelo. Multitemporal MERIS images for land cover mapping at national scale: the case study of Portugal. *International Journal of Remote Sensing*.
- Chuvieco, E., 1996. *Fundamentos de Teledetección espacial*. 3ª Edición, Madrid: Ediciones Rialp, S.A.
- Cochran, W., 1977. *Sampling Techniques*, New York: Wiley.
- Congalton, R. e Green, K., 1999. *Assessing the accuracy of remotely sensed data: principles and practices*. Danvers: CRC Press.
- Costa, H., Araújo, A., Carrão, H. e Caetano, M., 2008. Influência das características técnicas das imagens de satélite na produção de cartografia de ocupação do solo: estudo baseado em imagens MERIS e AWiFS, [CD-ROM]. In *Actas do X Encontro de Utilizadores de Informação Geográfica (ESIG 2008)*, Oeiras, Portugal.

- Dean, A. e Smith, G., 2003. An evaluation of per-parcel land cover mapping using maximum likelihood class probabilities. *International Journal of Remote Sensing*, 24 (14), 2905-2920.
- DGOTDU, 2004. *Contributos para a Identificação e Caracterização da Paisagem em Portugal Continental*. Volumes I a V, Lisboa: Direcção-Geral do Ordenamento do Território e Desenvolvimento urbano.
- Dicks, S. e Lo, T., 1990. Evaluation of thematic map accuracy in a land-use and land-cover mapping program. *Photogrammetric Engineering & Remote Sensing*, 56, 1247-1252.
- El-Magd, E. e Tanton, T., 2003. Improvements in land use mapping for irrigated agriculture from satellite sensor data using a multi-stage maximum likelihood classification. *International Journal of Remote Sensing*, 24 (21), 4197-4206.
- Foody, G., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, 185-201.
- Foody, G. e Mathur, A., 2004. A relative Evaluation of Multiclass Image Classification by Support Vector Machines. *IEEE transactions on Geoscience and Remote Sensing*, 42 (6), 1335-1343.
- Franklin, J., Woodcock, C. e Warbington, R., 2000. Multi-attribute vegetation maps of Forest Service lands in California supporting resource management decisions. *Photogrammetric Engineering & Remote Sensing*, 66, 1209-1217.
- Friedl, M. e Brodley, C., 1997. Decision tree classification of land Cover from remotely sensed data. *Remote Sensing of Environment*, 61, 399-409.
- Friedl, M., McIver, D., Hodges, J., Zhang, X., Muchoney, D., Strahler, A., Woodcock, C., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. e Schaaf, C., 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83, 287-302.
- Fuller, R., Groom, G. e Jones, A., 1994. The Land Cover Map of Great Britain: An automated Classification of Landsat Thematic Mapper Data. *Photogrammetric engineering & Remote Sensing*, 60 (5), 553-562.
- Gamanya, R., De Maeyer, P. e De Dapper, M., 2007. An automated satellite image classification design using object-oriented segmentation algorithms: A move towards standardization. *Expert Systems with Applications*, 32, 616-624.
- Gregorio, A. e Jansen, L., 1996. *FAO Land Cover Classification: A Dichotomous, Modular-Hierarchical Approach*. Food and Agriculture Organization, Sustainable Development Department (URL: <http://www.fao.org/sd/Eidirect/Eire0019.htm>, consulta a 28/07/2008).
- Hagner, O. e Reese, H., 2007. A method for calibrated maximum likelihood classification of forest types. *Remote Sensing of Environment*, 110, 438-444.
- Homer, C., Huang, C., Yang, L., Wylie, B., e Coan, M., 2004. Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric engineering & Remote Sensing*, 70 (7), 829-840.
- Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, H., McKerrow, A., VanDriel, J. e Wickham, J., 2007. Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric engineering & Remote Sensing*, 73 (4), 337-341.
- Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, IT-14 (1), 55-63.
- Jensen, R., 1996. *Introductory digital image processing: a remote sensing perspective*. 2ª edição, New Jersey: Prentice-Hall.
- Justice, C., Vermote, E., Townshend, J., Defries, R., Roy, D., Hall, D., Salomonson, V., Privette, J., Riggs, G., Strahler, A., Lucht, W., Myneni, R., Knyazikhin, Y., Running, S.,

- Nemani, R., Wan, Z., Huete, A., Leeuwen, W., Wolfe, R., Giglio, L., Muller, L., Lewis, P. e Barnsley, M., 1998. The moderate resolution imaging spectroradiometer (MODIS): land Remote Sensing for global change research. *IEEE transactions on Geoscience and Remote Sensing*, 36 (4), 1228-1249.
- Laba M., Gregory, S., Braden, J., Ogurcak, D., Hill, E., Fegraus, E., Fiore, F. e DeGloria S., 2002. Conventional and fuzzy accuracy assessment of the New York Gap Analysis Project land cover map. *Remote Sensing of Environment*, 81, 443-455.
- Laba, M., Downs, R., Smith, S., Welsh, S., Neider, C., White, S., Richmond, M., Philpot, W., Baveye, P., 2008. Mapping invasive wetland plants in the Hudson River National Estuarine Research Reserve using quickbird satellite imagery. *Remote Sensing of Environment*, 112, 286-300.
- Labovitz, M. 1986. Issues arising from sampling designs and band selection in discriminating ground reference attributes using remotely sensed data. *Photogrammetric Engineering & Remote Sensing*, 52 (2), 201-211.
- Lillesand, T., e Kiefer, R., 1994. *Remote sensing and image interpretation*. 3ª edição, New York: John Wiley & Sons, Inc.
- Loveland, T. e Belward, A., 1997. The IGBP-DIS global 1 km land cover data set, DISCover: first results. *International Journal of Remote Sensing*, 18 (15), 3289-3295.
- Mather, P. 2004. *Computer processing of remotely-sensed images: an introduction*, 3ª edição, Chippingham: Wiley.
- Montanarella, L., Rompaey, A. e Jones, R., 2003. Soil erosion risk in Europe. In *Proceedings of OECD expert meeting on soil erosion and soil biodiversity*, Roma, Italy, 87-100.
- NRSA, 2005. *National land use and land cover mapping using Multitemporal AWiFS data (LULC-AWiFS): project manual*. National Remote Sensing Agency, Department of Space, Government of India (URL [http://218.248.0.130:15001/images/lulc\\_manual.pdf](http://218.248.0.130:15001/images/lulc_manual.pdf), consulta em 25/10/2008).
- Omernik, J., 1987. Ecoregions of the Conterminous United States, Map (scale 1:7500000), *Annals of the Association of American Geographers* 77(1), 118-125.
- Pal, M. e Mather, P., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86, 554-565.
- Rogan, J., Franklin, J. e Roberts, A. 2002. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. *Remote Sensing of Environment*, 80, 143-156.
- Running, S., Loveland, T., e Pierce, L., 1994. A vegetation classification logic based on remote sensing for use in global biogeochemical models. *Ambio*, 23, 77- 81.
- Sabins, F., 1987. *Remote sensing principles and interpretation*, 2ª edição, New York: Remote Sensing Enterprises, Inc.
- Sesnie, S., Gessler, P., Finegan, B. e Thessler, S. 2008. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sensing of Environment*, 112, 2145-2159.
- Shalaby, A. e Tateishi, R., 2007. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. *Applied Geography*, 27, 28-41.
- Stehman, S., 2001. Statistical rigour and practical utility in thematic map accuracy, *Photogrammetric Engineering & Remote Sensing*, 67, 727-734.
- Stehman, S. e Czaplewski, R., 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment*, 64, 331-344.

- Vogelmann, J., Howard, S., Yang, L., Larson, C., Wylie, B., e Van Driel, J., 2001. Completion of the 1990's National Land Cover Data Set for the conterminous United States, *Photogrammetric Engineering & Remote Sensing*, 67, 650-662.
- Woodcock E., Collins, J., Gopal, S., Jakabhazy, V., Li, X., Macomber, S., Ryherd, S., Harward, V., Levitan, J.e Wu, Y., 1994. Mapping forest vegetation using Landsat TM imagery and a canopy reflectance. *Remote Sensing of Environment*, 50, 240-254.
- Wulder, M., Franklin, S., White, J., Linke, J. e Magnussen, S., 2006. An accuracy assessment framework for large-area land cover classification products derived from medium-resolution satellite data. *International Journal of Remote Sensing*, 27 (4), 663-683.
- Xu, M., Watanachaturaporn, P., Varchney, P., Arora, M., 2002. Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, 97, 322-336.

## **ANEXOS**

## 1. Imagens AWiFS (composição colorida 342)

Imagens de Primavera



Abril de 2006

Imagens de Verão

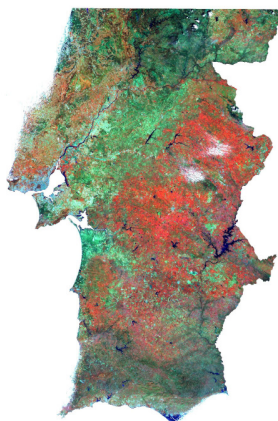


Julho de 2006

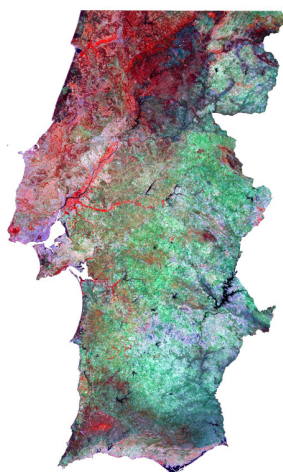
Imagens de Outono



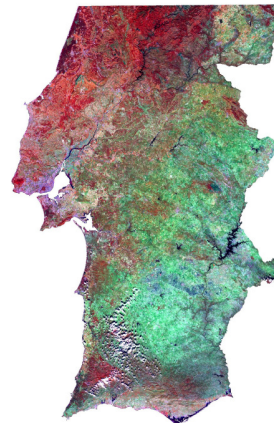
Outubro de 2006



Abril de 2006



Julho de 2006



Outubro de 2006

## 2. Nomenclaturas de ocupação do solo dos programas operacionais de cartografia de ocupação do solo

Suburban / rural development	Bracken
Urban development	Rough pasture / grass heath
Sea / estuary	Grass moor
Inland water	Lowland bog (herbaceous)
Beach / mudflats / cliffs	Upland bog (herbaceous)
Saltmarsh / intertidal vegetation	Open shrub heath
Inland bare ground	Open shrub moor
Tilled land (arable crops)	Dense shrub heath
Mown turf / grazed pasture	Dense shrub moor
Meadow / verge	Scrub / orchard
Rough / marsh grass	Deciduous / mixed woodland
Ruderal weed	Coniferous / evergreen woodland
Felled forest	

Tabela 2.1 – Nomenclatura do Mapa de ocupação do solo da Grã-Bretanha

Classes florestais do CLC2000	Classes florestais do Mapa de Ocupação do Solo	Classes florestais identificadas na classificação automática
Floresta de folhosas	Folhosas	Cortes florestais
Floresta de resinosas	Folhosas em zonas húmidas	Novas plantações
Floresta mista	Folhosas em rocha nua	Floresta jovem de coníferas
Áreas degradadas, cortes e novas plantações	Coníferas 5–15 m de altura	Floresta jovem de folhosas
	Coníferas >15 m de altura	Floresta de coníferas com altura inferior a 15m
	Coníferas com líquenes no sobcoberto	Floresta de coníferas com altura superior a 15m
	Coníferas em zonas húmidas	Floresta de coníferas dominadas por líquenes
	Folhosas em rocha nua	Floresta de folhosas
	Floresta mista	Floresta mista
	Floresta mista em zonas húmidas	
	Folhosas em rocha nua	
	Floresta jovem	
	Cortes de novas plantações	

Tabela 2.2 – Classes florestais do Mapa de Ocupação do Solo sueco e usadas na classificação automática



Nature conservation	Dryland horticulture
Other protected areas inc Indigenous uses	Irrigated pastures and cropping
Other minimal use	Irrigated horticulture
Grazing natural vegetation	Intensive animal and plant production
Production forestry	Rural residential
Plantation forestry	Urban intensive uses
Grazing modified pastures	Mining and waste
Dryland cropping	Water

Tabela 2.3 – Nomenclatura do mapa do programa *Australian Collaborative Land Use Mapping*

Open Water	Lichens
Perennial Ice / Snow	Moss
Developed, Open Space	Pasture / Hay
Developed, Low Intensity	Cultivated Crops
Developed, Medium Intensity	Woody Wetlands:
Developed, High Intensity	Palustrine Forested Wetland
Barren Land	Palustrine Scrub/Shrub Wetland
Unconsolidated Shore	Estuarine Forested Wetland
Deciduous Forest	Estuarine Scrub/Shrub Wetland
Evergreen Forest	Emergent Herbaceous Wetlands:
Mixed Forest	Palustrine Emergent Wetland (Persistent)
Dwarf Scrub	Estuarine Emergent Wetland
Shrub/Scrub	Palustrine Aquatic Bed
Grassland / Herbaceous	Estuarine Aquatic Bed
Sedge / Herbaceous	

Tabela 2.4 – Nomenclatura do projecto NLCD 2001

Build up land (urban / rural)	Wastelands:
Agriculture:	Other waste lands
Kharif crop land	Gullied / ravines
Rabi crop land	Land with shrub / scrub
Zaid crop land	Land without shrub / scrub
Double crop land	Shifting cultivated areas
Current fallow land	Others:
Plantations / Orchards	Snow covered / glacial areas
Forest:	Water bodies
Ever green / Semi ever green forest	
Deciduous forest	
Shrub / Degraded / Scrub forest	
Littoral swamp / Mangrove / Fresh water swamp	
Grassland & Grazing land	

Tabela 2.5 – Nomenclatura do Mapa de ocupação do solo do programa indiano NR-Census

### 3. Nomenclatura CORINE Land Cover

Nível 1	Nível 2	Nível 3
1. Territórios artificializados	1.1 Tecido urbano	1.1.1 Tecido urbano contínuo
		1.1.2 Tecido urbano descontínuo
	1.2 Indústria, comércio e transportes	1.2.1 Indústria, comércio e equipamentos gerais
		1.2.1 Redes viárias e ferroviárias e espaços associados
		1.2.3 Zonas portuárias
		1.2.4 Aeroportos
	1.3 Áreas em construção, de extracção e de depósitos de resíduos	1.3.1 Áreas de extracção mineira
		1.3.2 Áreas de deposição de resíduos
		1.3.3 Áreas em construção
	1.4 Zonas verdes ordenadas	1.4.1 Espaços verdes urbanos
		1.4.2 Equipamentos desportivos e de lazer
2. Áreas agrícolas	2.1 Culturas anuais	2.1.1 Culturas anuais de sequeiro
		2.1.2 Culturas anuais de regadio
		2.1.3 Arrozais
	2.2 Culturas permanentes	2.2.1 Vinhas
		2.2.2 Pomares
		2.2.3 Olivais
	2.3 Pastagens	2.3.1 Pastagens
	2.4 Áreas agrícolas heterogéneas	2.4.1 Culturas anuais associadas às culturas permanentes
		2.4.2 Sistemas culturais e parcelares complexos
		2.4.3 Agricultura com espaços naturais
		2.4.4 Sistemas agro-florestais
3. Florestas e meios semi-naturais	3.1 Florestas	3.1.1 Floresta de folhosas
		3.1.2 Floresta de resinosas
		3.1.3 Florestas mistas
	3.2 Vegetação arbustiva e herbácea	3.2.1 Pastagens naturais
		3.2.2 Matos
		3.2.3 Vegetação esclerófitica
		3.2.4 Espaços florestais degradados, cortes e novas plantações
	3.3 Zonas descobertas e com pouca vegetação	3.3.1 Praias, dunas e areais
		3.3.2 Rocha nua
		3.3.3 Vegetação esparsa
		3.3.4 Áreas ardidas
		3.3.5 Neves eternas e glaciares
4. Zonas húmidas	4.1 Zonas húmidas interiores	4.1.1 Pauis
		4.1.2 Turfeiras
	4.2 Zonas húmidas costeiras	4.2.1 Sapais
		4.2.2 Salinas
		4.2.3 Zonas interditas
5. Massas de água	5.1 Águas interiores	5.1.1 Linhas de água
		5.1.2 Planos de água
	5.2 Águas marinhas	5.2.1 Lagunas litorais
		5.2.2 Estuários
		5.2.3 Mar e oceano

Tabela 3.1 – Nomenclatura do CORINE Land Cover

#### 4. Nomenclatura LANDEO e nomenclatura adoptada

LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
1. Artificial Areas	1.1 Continuous Artificial Areas	1.1.0.	1.1.0.0.
	1.2 Descontinuous Artificial Areas	1.2.0.	1.2.0.0.
	2.1. Rainfed Herbaceous Crops	2.1.0.	2.1.0.0.
	2.2. Irrigated Herbaceous Crops	2.2.0.	2.2.0.0.
	2.3. Rice Crops	2.3.0.	2.3.0.0.
2. Croplands	2.4. Permanent Crops (Trees or Shrubs)	2.4.1. Permanent Evergreen Crops (Trees or Shrubs)	2.4.1.1. Permanent Evergreen Crops (Trees or Shrubs) without understory agricultural systems 2.4.1.2. Permanent Evergreen Crops (Trees or Shrubs) with understory agricultural systems
		2.4.2. Permanent Deciduous Crops (Trees or Shrubs)	2.4.2.1. Permanent Deciduous Crops (Trees or Shrubs) without understory agricultural systems 2.4.2.2. Permanent Deciduous Crops (Trees or Shrubs) with understory agricultural systems
	2.5. Mosaic Cultivated and managed lands	2.5.0..	2.5.0.0.
3. Natural And Semi-Natural Terrestrial Vegetation	3.1. Broadleaved Closed to Open Trees	3.1.1. Broadleaved Closed Trees	3.1.1.1 Broadleaved Closed Trees without understory agricultural systems 3.1.1.2 Broadleaved Closed Trees with understory agricultural systems
			3.1.2.1 Broadleaved Open Trees without understory agricultural systems 3.1.2.2. Broadleaved Open Trees with understory agricultural systems
		3.1.2. Broadleaved Open Trees	
	3.2. Needleleaved Closed to Open Trees	3.2.1. Needleleaved Closed Trees	3.2.1.0.
		3.2.2. Needleleaved Open Trees	3.2.2.0.
	3.3. Mixed Closed to Open Trees	3.3.1. Mixed Closed Trees	3.3.1.0.
		3.3.2. Mixed Open Trees	3.3.2.0.
	3.4. Closed to Open Shrubland	3.4.1. Closed Shrubland	3.4.1.0.
		3.4.2. Open Shrubland	3.4.2.0.
	3.5. Closed to Open Herbaceous Vegetation	3.5.1 Closed Herbaceous Vegetation	3.5.1.0.
		3.5.2 Open Herbaceous Vegetation	3.5.2.0.
	3.6. Sparse Vegetation	3.6.1 Sparse Trees	3.6.1.0.
		3.6.2 Sparse Shrubs	3.6.2.0.
		3.6.3 Sparse Herbaceous	3.6.3.0.
4. Mosaics	3.7. Mosaic Trees/Shrubs/Herbaceous	3.7.0.	3.7.0.0.
	3.8. Recently Burnt (Trees or Shrubs)	3.8.0	3.8.0.0.
	3.9. Clear-cutts	3.9.0	3.9.0.0.
	3.10. New plantations	3.10.0	3.10.0.0.
	4.1. Mosaic Cultivated and managed lands/Natural and Semi-natural areas/Artificial areas	4.1.0	4.1.0.0.
5. Permanent Wetlands	5.0	5.0.0	5.0.0.0.
6. Barren	6.0	6.0.0	6.0.0.0.
7. Water Bodies	7.0	7.0.0	7.0.0.0.

Tabela 4.1 – Nomenclatura LANDEO

Código	Classe de ocupação do solo	Descrição
11 (1)	Continuous Artificial Areas (Áreas artificiais contínuas)	A ocupação do solo consiste em áreas artificiais (e.g., edificações, estradas). Pelo menos 80% da área total deverá ser impermeável.
12 (2)	Discontinuous Artificial Areas (Áreas artificiais descontínuas)	A ocupação do solo consiste em áreas artificiais (e.g., edificações, estradas). Entre 30-80% da área total deverá ser impermeável.
21 (3)	Rainfed Herbaceous Crops (Culturas de sequeiro)	A ocupação do solo consiste em culturas herbáceas de sequeiro. As culturas são colhidas anualmente e seguidas por um período de solo nu.
22 (4)	Irrigated Herbaceous Crops (Culturas de regadio)	A ocupação do solo consiste em culturas herbáceas de regadio. As culturas são colhidas anualmente e seguidas por um período de solo nu.
23 (5)	Rice Crops (Arrozais)	A ocupação do solo consiste em arrozais. As culturas são colhidas anualmente e seguidas por um período de solo nu.
242 (6)	Permanent Deciduous Crops (Trees or Shrubs) (Vinhas)	A ocupação do solo consiste em culturas permanentes de folha caduca. Inclui vinhas e outras plantações.
311 (7)	Broadleaved Closed Trees (Floresta de folhosas)	A ocupação do solo consiste em árvores de espécies de folhosas com pelo menos 5m de altura e com uma cobertura do solo superior a 40%.
312 (9)	Broadleaved Open Trees (Áreas agro-florestais)	A ocupação do solo consiste em árvores de espécies de folhosas com pelo menos 5m de altura e com uma cobertura do solo entre 15-40%.
321 (8)	Needleleaved Closed Trees (Floresta de resinosas)	A ocupação do solo consiste em árvores de espécies de resinosas com pelo menos 5m de altura e com uma cobertura do solo superior a 40%.
34 (10)	Closed to Open Shrubland (Matos)	A ocupação do solo consiste em vegetação lenhosa (arbustos) que cobre mais que 15% do solo e tem menos de 5m de altura. A cobertura arbórea é inferior a 15%.
35 (11)	Closed to Open Herbaceous Vegetation (Vegetação herbácea natural)	A ocupação do solo consiste em vegetação herbácea que cobre mais que 15% do solo. A cobertura arbórea e arbustiva é inferior a 15%.
39 (12)	Clear-cuts (Cortes florestais)	A ocupação do solo consiste na substituição de formações florestais de resinosas ou folhosas por solo nu devido a um corte florestal executado durante o ano de referência da cartografia
5 (13)	Permanent Wetlands (Zonas húmidas)	A ocupação do solo consiste numa combinação permanente de água e vegetação. A vegetação pode estar presente em água salgada, salobra ou doce.
6 (14)	Barren (Solo nu)	A ocupação do solo consiste em áreas com menos de 4% de cobertura vegetal durante todo o ano (e.g. rocha nua, areais).
7 (15)	Water Bodies (Água)	A ocupação do solo consiste em corpos de água naturais ou artificiais de água doce ou salgada

Tabela 4.2 – Classes LANDEO adoptadas na presente dissertação (correspondentes códigos e nomes das classes da nomenclatura entre parêntesis) e sua descrição baseada no LCCS

## 5. Exactidão global e classes de cada fase de classificação do teste 15

Fase de classificação	Número de classes	Classes de ocupação do solo	Exactidão global (%)
1	2	Água e zonas húmidas Classes terrestres	99.6
2	4	Água Zonas húmidas Classes sem vegetação Classes com vegetação	94.4
3	8	Água Zonas húmidas Áreas artificiais contínuas Áreas artificiais descontínuas Solo nu Classes com vegetação não agrícola e não herbácea Áreas agrícolas e vegetação herbácea natural	79.6
4	12	Água Zonas húmidas Áreas artificiais contínuas Áreas artificiais descontínuas Solo nu Cortes florestais Área agro-florestal Matos Culturas de sequeiro e vegetação herbácea natural Vinha Floresta Culturas de verão	69.6
5	15	Água Zonas húmidas Áreas artificiais contínuas Áreas artificiais descontínuas Solo nu Cortes florestais Área agro-florestal Matos Culturas de sequeiro vegetação herbácea natural Vinha Floresta de folhosas Floresta de resinosas Culturas de regadio Arrozais	60.9

Tabela 5.1 – Exactidão global e classes de cada fase de classificação do teste 15

## 6. Comparação gráfica entre os melhores mapas obtidos com o CMV e AD

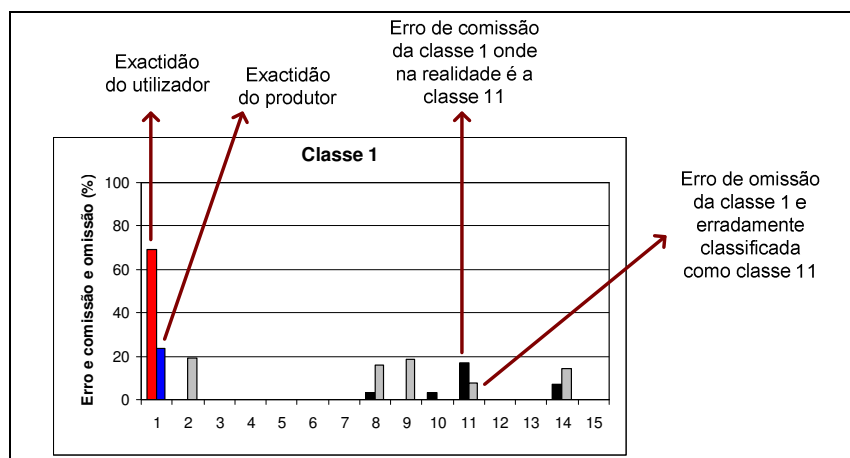
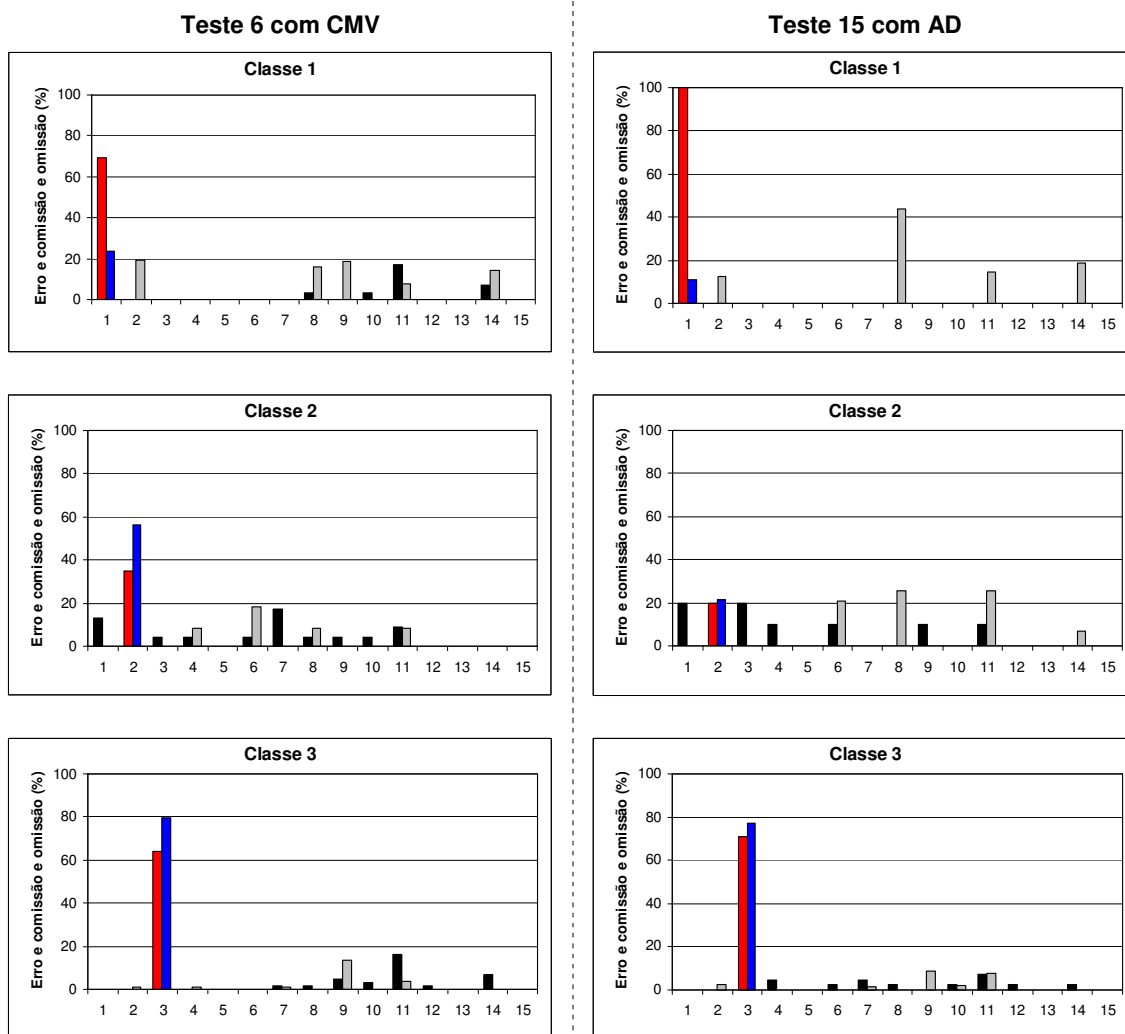
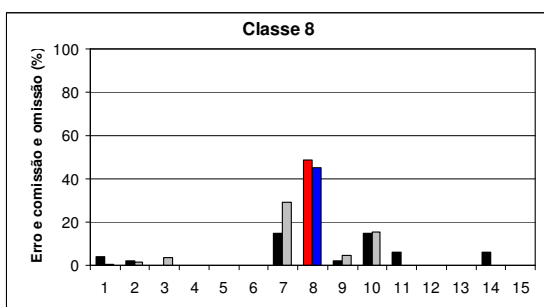
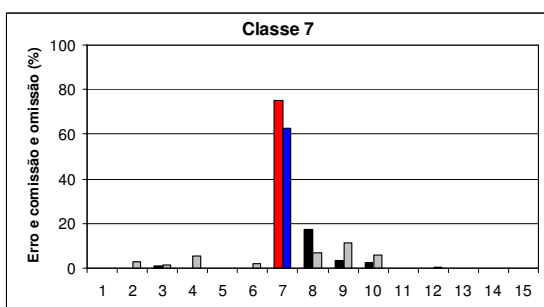
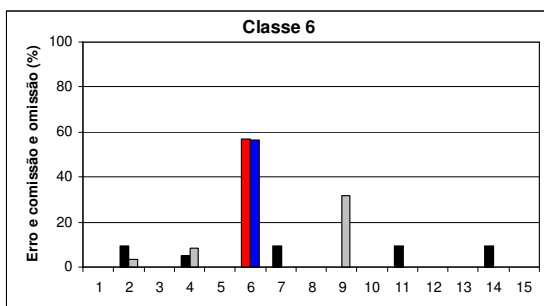
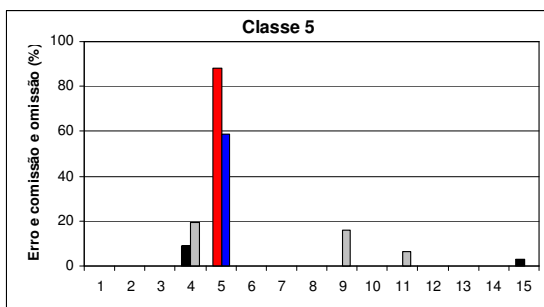
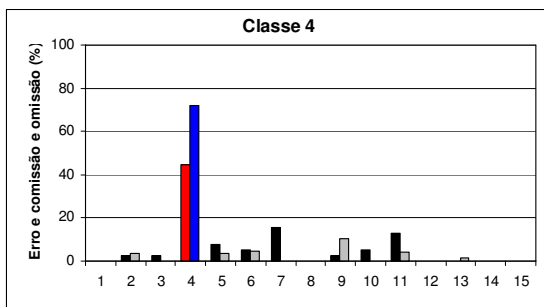


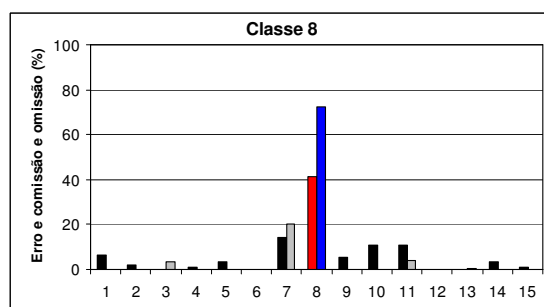
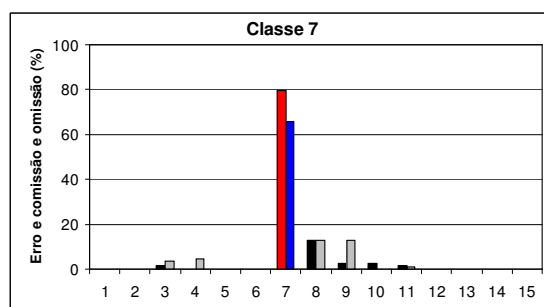
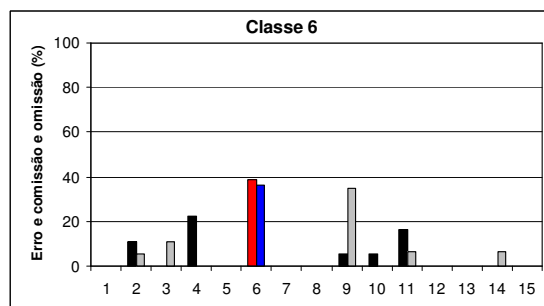
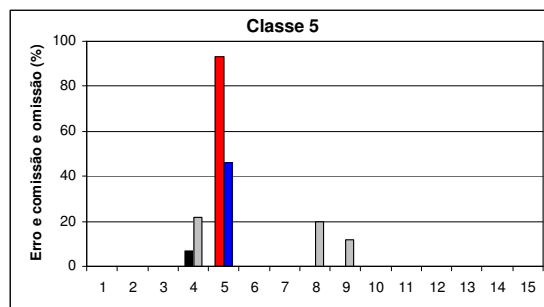
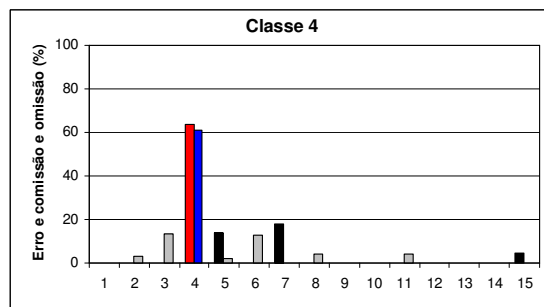
Figura 6.1 – Exemplo de interpretação dos gráficos do Anexo 6 (classe 1 com o CMV do teste 6)



### Teste 6 com CMV

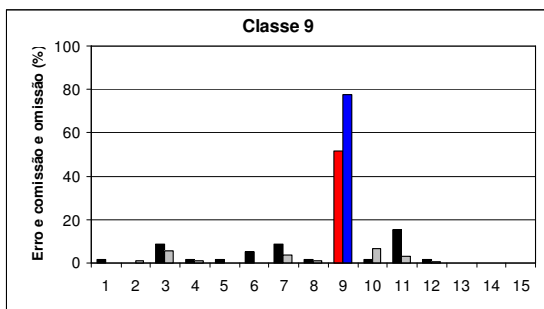


### Teste 15 com AD

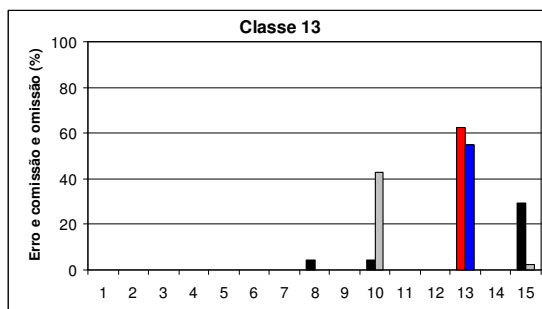
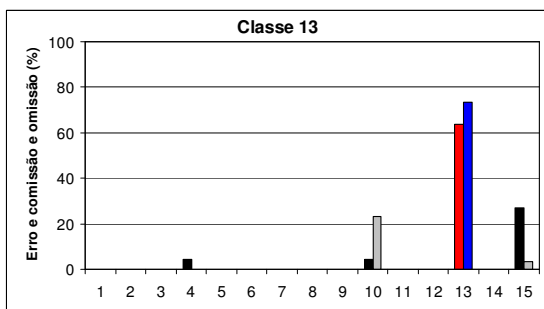
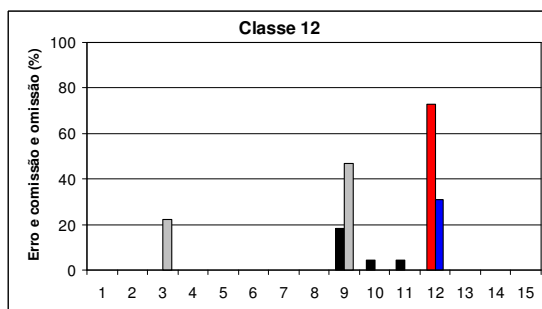
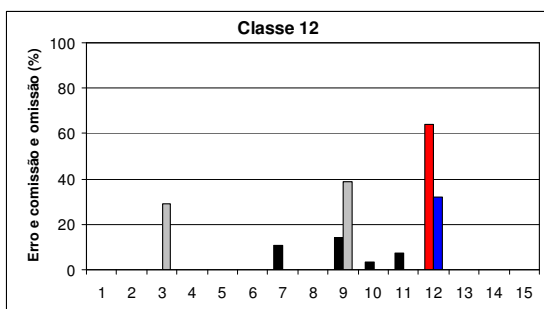
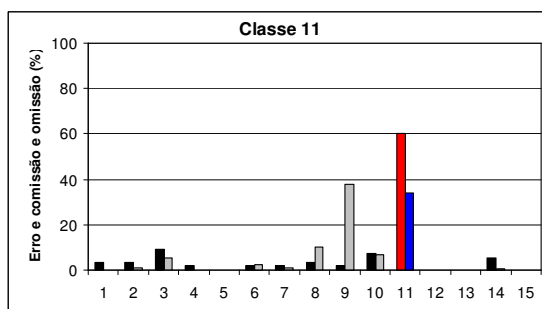
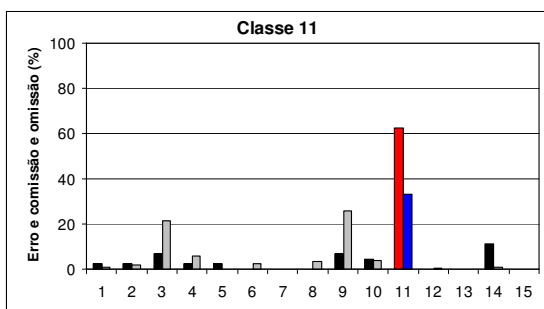
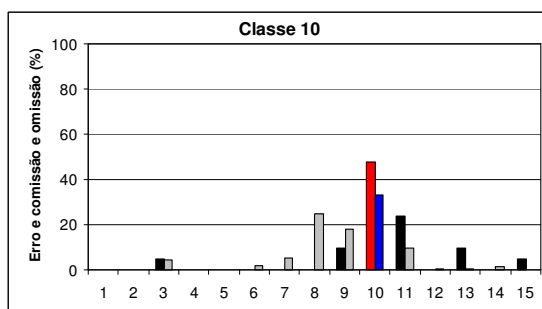
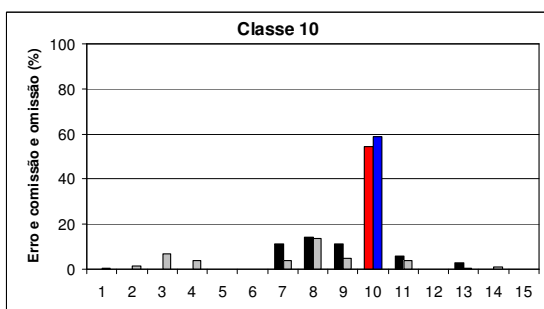
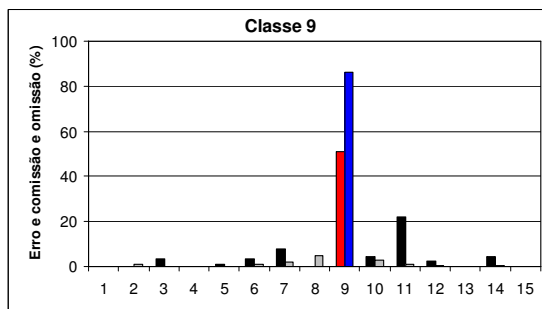




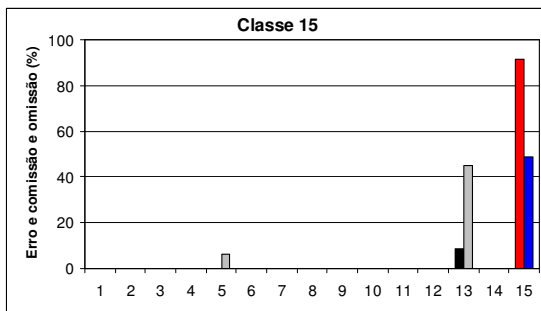
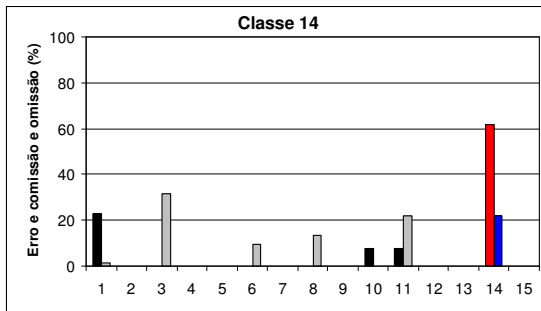
Teste 6 com CMV



Teste 15 com AD



**Teste 6 com CMV**



**Teste 15 com AD**

